

Result No.	Query			ID	Description
	Score	Match	Length		
1	287.5	11.5	730	PHLC_PSEAE	P06200 pseudomonas
2	267	10.7	692	PHLN_PSEAE	P15713 pseudomonas
3	167.5	6.7	421	PHOX_KLUJA	P08540 kluyveromyc
4	119.5	4.8	417	PHOA_ASPNG	P34724 aspergillus
5	116	4.6	412	PHOA_PENCH	P37272 penicillium
6	112.5	4.5	488	PHB_ALCPA	P12625 alcaligenes
7	110.5	4.4	729	COAL_PAVP9	P33484 porcine par
8	109.5	4.4	729	COAL_PAVP8	P32501 porcine par
9	107	4.3	1257	CCAA_BACTV	P45734 bacillus th
10	106.5	4.3	1341	YL78_YEAST	P05894 saccharomyc
11	105.5	4.2	1211	ATHL_YEAST	P48016 saccharomyc
12	105.5	4.2	4590	FATH_HUMAN	Q14517 homo sapien
13	105	4.2	749	COA2_PAVPN	P22964 porcine par
14	104.5	4.2	625	BGAL_LACSK	Q48846 lactobacill
15	104.5	4.2	729	COAL_PAVPN	P18546 porcine par
16	104.5	4.2	991	PCGI_PORGI	P28784 porphyrom
17	104.5	4.2	1122	ADPL_MYCGA	Q49379 mycoplasma
18	101.5	4.1	800	INUA_LISMO	P35146 listeria mo
19	101	4.0	1146	KTAA_KLUJA	P09805 kluyveromyc
20	100.5	4.0	736	CPG2_PORGI	P95493 porphyrom
21	99	4.0	434	GLN2_HORVU	P13564 hordeum vul
22	99	4.0	1645	OMPB_RICTY	P96989 r outer mem
23	98	3.9	430	GLN2_PEA	P08281 pisum sativ
24	97.5	3.9	458	YM51_YEAST	P422933 saccharomyc
25	97	3.9	558	P4HA_CAEEL	Q10576 caenorhabdi
26	97	3.9	584	D29B_ARATH	Q06737 arabidopsis
27	97	3.9	600	LT65_ARATH	Q04980 arabidopsis
28	96.5	3.9	775	VP4_ROTFI	P39034 feline rota
29	96.5	3.9	1006	RATL_YEAST	Q02752 saccharomyc
30	96.5	3.9	2148	VITI_AEDRA	Q16927 aedes aegypt
31	96.5	3.9	4447	PKSK_BACSU	P40803 bacillus su
32	96	3.8	532	GSI_NEUCR	P38678 neurospora
33	96	3.8	659	AMLA_STRPN	P18791 streptococc

QY	11	GLLASGLASPVTSVTSVREAPFGYKPGKSEIENLKDKNVENIWMILLIENSFONILG- 69
Db	15	GTVAATVATSNLPSSTQALAIPAHRHG-----NLKD-VEHVVILMOENSPDHYGT 67
QY	70	--GYRRG--LDNPNINGPCNKYKASDPSSCKY-----CTQAKDYDVFNPDHSHVTGN 120
Db	68	LKGYRGFGDMAIPLDQGRVWHQKSGKEILLPYHFTSTSAQRVDG----- 115
QY	121	NLEFYGYTTPNGATASGVYADQSGFLNAQLNDPKLAPAEATRVGMGYTTEEVPTLV 180
Db	116	-----TPHTWPA-----QAANEGRMD--KWLPAK-TERSLGYKKEODIAFOF 156
QY	181	DLVDEFTFNFSFCVPGTPNPSELCAAGT-----AAGHG-----KNDDF---LNYGTSK 230
Db	157	AMANAFTICDAYHCFSQGGPNPLFLWTGTNDPLQGHGGVTTNHDHSGNPGVEQGTWT 216
QY	231	SIFEAANEKGVSLMYDGTNGEFEPDSL---FFTYVNOTSRNV-----PVENFTODA 281
Db	217	TYPERLQAAGITWRVYQDMADNFDNPLGFRQYRAAAPDSPLVNGLSTWNKLDAKRDV 276
QY	282	YLGVLPRKFINPSCGTTNTNSMHP-TGNVSYGEVFKQYDAIROGPO-WDKTLFFITY 339
Db	277	LANSLPOVSWI-----VAPAKYSBHPGSSPIWGAETYSWLDALTANPEWMSKTAALLVMF 332
QY	340	DETGGEYDVHVPPLA--VRPNLTYTET-----AKNGQKYTLHFDRGLGRMPTWVI 368
Db	333	DENDGDFHVAAPAPSLNKGDTLRCKTTADATLEWHTKGDIBRYNQPGVGLGARVPMTVI 392
QY	389	SPYSKGYTEQGTDPVTGKAPYSATSVLTKL 421
Db	393	SPNSKGGWNS-----QVFDHTSVIREL 415

```

RESULT 3
PHOX_KLUJA
ID PHOX_KLUJA STANDARD; PRT; 421 AA.
AC P08540;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE POTENTIAL ACID PHOSPHATASE (EC 3.1.3.2).
QS Kluveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Kluveromyces.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-CBS 2359 / IFO 1267 / NRRL Y-1140;
RX MEDLINE-89034156; PubMed-3053697;
RA Chang Y.-D., Dickson R.C.;
RT "Primary structure of the lactose permease gene from the yeast
RL Kluveromyces lactis. Presence of an unusual transcript structure.";
RL J. Biol. Chem. 263:16696-16703(1988).
CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN
CC ALCOHOL + ORTHOPHOSPHATE.
CC -!- SIMILARITY: VERY STRONG, TO ASPERGILLUS NIGER AND PENICILLIUM
CC CHRYSOGENIUM PHOA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06937; CAA30054.1; ALT_INIT.
CC PIR; B31776; B31776.
CC Hypothetical protein; Hydrolase.
KW

```



FT ACT\_SITE 228 228 PROTON DONOR (BY SIMILARITY).  
 SQ SEQUENCE 421 AA; 47193 MW; 2F7E614C943824E6 CRC64;  
  
 Query Match 6.7%; Score 167.5; DB 1; Length 421;  
 Best Local Similarity 21.0%; Pred. No. 7.5e-05;  
 Matches 89; Conservative 65; Mismatches 147; Indels 123; Gaps 23;  
  
 QY 122 LEYGYTPNNGATAS-----GKVADQSG-----FLNAQLNDYPKLAPEEATROV 167  
 DB 31 LRTYTSISLSEIESAASATEVAEVSVDGGAFAKRFETIFLENTDYDKAAGDES----- 86  
 QY 168 MGYTTEEVPTLVLDVEFTFNWFCVPGPTNPNRLCALAGTAAGHGKNDGDFLNYGI 227  
 DB 87 LSLAEOGI-----TLTNW--ALTHPSEPNLASVGDI--FALDDDFRFSMPS 132  
 QY 228 SSKSIFEAANEKGYSLNVD-----GTNG-----EEPDLSLRTYVNOT 266  
 DB 133 NVSNIVDLDLTKGISMAEYQEHSPAGFGMNFNSQETVASYVVKHNPILLFDNVVND 192  
 QY 267 SR-SNNVVPVENFFODAYGLVLPKFSYINPSCCGTNTSMHPTGNVSYGEVEVKQIYDAIR 325  
 DB 193 TRLANIKNFEDFNVDENEKLPYAFITPNM-----TNDGHD-TIQFAGKWSKDFLAPLL 247  
 QY 326 QGPQW-DKTLFIYDET-----GG-----FYDHVP----- 350  
 DB 248 ENDYFMDTLLVLLTFDENETYGKKNVFSILLGGVIEDELKTKDQDTFYDHSOLASVEA 307  
 QY 351 ----PPLAVRPDLNLTYYTAKNGQYT-LHFDRLGGRMPTWISPSYKGYEYQYDGPV 405  
 DB 308 NWDPLHLGRHGDGANVLEIVANATNITNVEVD-----TMYMINE-TYIGYLDNYIE-- 358  
 QY 406 TGPAP-----YSATSLATLGLYLDIEDTPRAHSPSFEDHLIGTLREDAPIALKTP 459  
 DB 359 --LPAPWTAIRNQCIPILDSIKETWE-DEYSKVQSES-----YYSTTTTYSADVT--DA 409  
 QY 460 HTFS 463  
 DB 410 ETFS 413  
  
 RESULT 4  
 PHOA\_ASPNG  
 ID PHOA\_ASPNG STANDARD; PRT; 417 AA.  
 AC P34724;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2).  
 GN PHOA.  
 OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;  
 CC anamorphic Trichocomaceae; Aspergillus.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-NRRL 3135 / VAN TIEGHEM / FICUUM;  
 RA MEDLINE=95032134; PubMed=7945393;  
 RX Ehrlich K.C., Montalbano B.G., Mullaney E.J., Dischinger H.C. Jr.,  
 RA Ullah A.H.J.;  
 RT "An acid phosphatase from Aspergillus ficuum has homology to  
 RT Penicillium chrysogenum Phoa.",  
 RL Biochem. Biophys. Res. Commun. 204:63-68(1994).  
 CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN  
 CC ALCOHOL + ORTHOPHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -!- SIMILARITY: TO P-CHRYSOGENUM PHOA AND TO A KLUYVEROMYCES LACTIS  
 CC ORF (AC P08540).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 CC EMBL; L20566; AAA62393.1; .  
 CC Hydrolase; Glycoprotein; Signal.  
 CC SIGNAL 1 19  
 FT CHAIN 20 417  
 FT ACT\_SITE 215 215 PROTON DONOR (BY SIMILARITY).  
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 417 AA; 46735 MW; 896FD2AF541C69C8 CRC64;  
  
 Query Match 4.8%; Score 119.5; DB 1; Length 417;  
 Best Local Similarity 20.6%; Pred. No. 0.19;  
 Matches 74; Conservative 43; Mismatches 137; Indels 105; Gaps 15;  
  
 QY 155 YPKLAPEEATROVMGYT-----EVEVPTLVYD---LV 183  
 DB 26 YPSLAETRAAQATVQPSVSNVKGLTFRNFVNIWLENTDFDAAAATEHLPLAKKGLLL 85  
 QY 184 DEFTTFNSWFCVPGPTNPNRLCALAGTAAGHGKNDGDFLNYGSSKSIFFAANEKGVSW 243  
 DB 86 NNF-----W--AVTHPSEPN-YCRHLPLGDTFGMDNDDFHQPSNVSTIALDFTKNIAW 137  
 QY 244 -----LNVDTNGEPEFDS-----LFTYVNTQTSR-SNNVVPVENFFQDAYL 283  
 DB 138 GEYQEGLPYGYQYRYPESGANDYVRNRNPLILFDSVTEADLRRLQKNFSSEYDLEN 197  
 QY 284 GVLPKFSYINPSCCGTNTSMHPTGNVSYGEVEVKQIYDAIRGQPDWK-TLLFTTYDET 342  
 DB 198 HRLPQWFIPTNM-----TNDGHD-TNIFSGDWTGWFSELLENDYFKDILMLTDEET 252  
 QY 343 GGFY-----DHVPPPLAVRPDLNLTYT-----ETAKNGQYITLHFDRLGGRMPT 385  
 DB 253 GTYEIGNNIYTFLLGGAVPDLLGTGKDDTFYTHYSVIASTLWGLPSLGRMDCGANLFS 312  
 QY 386 WVI-----SPYSKKGYTEOYGTDPDTGPKAPYSA-TSVLKLTL 421  
 DB 313 WLAKKTYGVNVEVDFSNLYNETHWGLPSLDDDDYSEYAGWPVPTTDDSCSAGNGILSTV 371  
  
 RESULT 5  
 PHOA\_PENCH  
 ID PHOA\_PENCH STANDARD; PRT; 412 AA.  
 AC P37274;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PHOSPHATE-REPRESSIBLE ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2).  
 GN PHOA.  
 OS Penicillium chrysogenum.  
 OC Eukaryota; Fungi; Ascomycota; anamorphic Ascomycota; Penicillium.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=92225342; PubMed=1563629;  
 RA Haas H., Redl B., Friedlin E., Stoeffler G.;  
 RT "Isolation and analysis of the Penicillium chrysogenum phoa gene  
 RT encoding a secreted phosphate-repressible acid phosphatase.",  
 RL Gene 113:129-133(1992).  
 CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN  
 CC ALCOHOL + ORTHOPHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -!- SIMILARITY: TO A NIGER PHOA AND TO A KLUYVEROMYCES LACTIS ORF  
 CC (AC P08540).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE COAT PROTEIN VP1 [CONTAINS: COAT PROTEIN VP2].  
 OS Porcine parvovirus (strain 90HS) (PPV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-89319168; PubMed-2750278;  
 RA Sakurai M., Nishimori T., Ushimi C., Nakajima H.;  
 RT "Nucleotide sequence of capsid protein gene of porcine parvovirus.";  
 RL Virus Res. 13:79-86(1989).  
 CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED  
 CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.  
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT  
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.  
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 DR PIR: A60006; A60006.  
 DR HSP: P30129; 4DPV.  
 DR INTERPRO: IPR001403; -.  
 DR PFAM: PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CHAIN 1 729  
 FT CARBOHYD 151 729  
 FT CARBOHYD 172 729  
 FT CARBOHYD 198 729  
 FT CARBOHYD 282 729  
 FT CARBOHYD 330 729  
 FT CARBOHYD 433 729  
 FT CARBOHYD 471 729  
 FT CARBOHYD 573 729  
 FT CARBOHYD 604 729  
 FT CARBOHYD 651 729  
 SQ SEQUENCE 729 AA; 80938 MW; 92538BFF9A0C78E6 CRC64;  
  
 Query Match 4.4%; Score 110.5; DB 1; Length 729;  
 Best Local Similarity 21.7%; Pred. No. 1.8;  
 Matches 88; Conservative 45; Mismatches 145; Indels 127; Gaps 23;  
  
 QY 110 FNDPDSHVTGNL-----EFYG-----TYTPNNGAIAAGKVVADQS 145  
 Db 271 FNPADQLISNNMTNINLVSEQIEFNVVLKTTESATSPPTKIYNNDLTASLVALDNTN 330  
 QY 146 GFNLQAOLNDYPKLAPPEATROVMGYTTEEVPTLVLDVDEFTTFNSWFCVPGPTNPRL 205  
 Db 331 NTL-----PYTPAARSET---LGFY---PWLPT-----KPTQYRYLSCTRNLP--- 371  
 QY 206 CALAGTAAGHKNDGDDFLNIGISKSIFEAANEKGVSWLNYDG---TNG--EFEPDSLFF 260  
 Db 477 A-----TAIRPAQGVYTPYMNFEYNSNGSPFLTPVPTA-----DTQNDDEPGAIRF 525  
 QY 315 VFVKQIYDAIROGP-QWDKTLFFITFYDETCGFYDHPVPPPLAVRPDLNLTETAKNGQ-KY 372  
 Db 477 A-----TAIRPAQGVYTPYMNFEYNSNGSPFLTPVPTA-----DTQNDDEPGAIRF 525  
 QY 373 TLHFDRLGGRMPT-----WVISPYSKGYI--BOYG-----TDQNDDEPGAIRF 525  
 Db 526 TMGYQH--GOLTTSSQELERYTFNPOSKGRAPQOQFNQOAPNLNENTNNGTLPLSDPIG 583  
 QY 407 GKP-----APYSA---TSVLKTLGYLWDIE---DFTPRV 434  
 Db 584 GKPNHFMNTLNTYGPLTALNTAPVPFNGQIWDKDLTDLPKRL 628  
  
 RESULT 8  
 COAL\_PAVPK  
 ID COAL\_PAVPK STANDARD; PRT: 729 AA.  
 AC P52501;  
 DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE COAT PROTEIN VP1 [CONTAINS: COAT PROTEIN VP2].  
 OS Porcine parvovirus (strain Kresse) (PPV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96183900; PubMed-8642680;  
 RA Bergeron J., Hebert B., Tijssen P.;  
 RT "Genome organization of the Kresse strain of porcine parvovirus;  
 RT identification of the allotropic determinant and comparison with  
 RT those of NADL-2 and field isolates.";  
 RL J. Virol. 70:2508-2515(1996).  
 CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED  
 CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.  
 CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT  
 CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.  
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 DR EMBL: U44978; AAC40230.1; -.  
 DR EMBL: U44978; AAC40231.1; -.  
 DR HSP: P30129; 4DPV.  
 DR INTERPRO: IPR001403; -.  
 DR PFAM: PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CHAIN 1 729  
 FT CARBOHYD 151 729  
 FT CARBOHYD 172 729  
 FT CARBOHYD 198 729  
 FT CARBOHYD 282 729  
 FT CARBOHYD 330 729  
 FT CARBOHYD 433 729  
 FT CARBOHYD 471 729  
 FT CARBOHYD 573 729  
 FT CARBOHYD 604 729  
 FT CARBOHYD 651 729  
 SQ SEQUENCE 729 AA; 80835 MW; B6345BFA0568A1F6 CRC64;  
  
 Query Match 4.4%; Score 109.5; DB 1; Length 729;  
 Best Local Similarity 21.7%; Pred. No. 2.1;  
 Matches 88; Conservative 45; Mismatches 145; Indels 127; Gaps 23;  
  
 QY 110 FNDPDSHVTGNL-----EFYG-----TYTPNNGAIAAGKVVADQS 145  
 Db 271 FNPADQLISNNMTNINLVSEQIEFNVVLKTTESATSPPTKIYNNDLTASLVALDNTN 330  
 QY 146 GFNLQAOLNDYPKLAPPEATROVMGYTTEEVPTLVLDVDEFTTFNSWFCVPGPTNPRL 205  
 Db 331 NTL-----PYTPAARSET---LGFY---PWLPT-----KPTQYRYLSCTRNLP--- 371  
 QY 206 CALAGTAAGHKNDGDDFLNIGISKSIFEAANEKGVSWLNYDG---TNG--EFEPDSLFF 260  
 Db 372 ----TYTGOSQOITDSITQGLHSDIMFYTIENAVPIHLLRTGDEFSTGIYHDFKPKL 426  
 QY 261 TYVNTQSRNVYPVENFEQ-----DAYLGLPKFSYINPSCCGTNT--NSMHPGTGNSVYGE 314  
 Db 427 THSQWTRSLGLPPKLLTEPTTEGQHPGTLF-----AANTRKGYHOTMNNYSYE 476  
 QY 315 VFVKQIYDAIROGP-QWDKTLFFITFYDETCGFYDHPVPPPLAVRPDLNLTETAKNGQ-KY 372  
 Db 477 A-----TAIRPAQGVYTPYMNFEYNSNGSPFLTPVPTA-----DTQNDDEPGAIRF 525  
 QY 373 TLHFDRLGGRMPT-----WVISPYSKGYI--BOYG-----TDQNDDEPGAIRF 525  
 Db 477 A-----TAIRPAQGVYTPYMNFEYNSNGSPFLTPVPTA-----DTQNDDEPGAIRF 525  
 QY 373 TLHFDRLGGRMPT-----WVISPYSKGYI--BOYG-----TDQNDDEPGAIRF 525  
 Db 477 A-----TAIRPAQGVYTPYMNFEYNSNGSPFLTPVPTA-----DTQNDDEPGAIRF 525







```

Db 4021 SNPCGVN-----PSPAGGYCKCSALY-----IGTHCEISNVPCCSNPCLY 4065
Qy 137 SKKVAGDSGFL-----NAQLNDYKPLAP-----EATQVNGYTYTEE 175
Db 4066 GGTGVNDGGFVOCGRGLYTGORCOLSPYCKDEPKNGGTGTFDSLGAOCODSGFRGR 4125
Qy 176 VPTLVLDVDEFTTFNSWFSCVPGPTNPNRLCALAGTAAGCHKN-----DDDFLNY 225
Db 4126 COSDID-----ECSGNPLCHGALC--ENTHSGYHCNCSEHYRGRHCEADAAPNQ 4171
Qy 226 GISSKSIPEANKEGYSWLNVDCTNGEEFDSLFYVYVQVTSRSNVVVENFODAYLGV 285
Db 4172 YVSTPMNGLAEGIGI-----VVFVAGIFLLVVV 4202
Qy 286 L-----PKFSYINPSCCGTNTSMHPTGNSYGEVFKQIYDAIRQGPQWD 331
Db 4203 LCRKMISRKKHQAEPKDKHLGPA-----TAFLQRPYFD 4236
Qy 332 KTLLEFYDETCGFYDHVPPPLAVRPNLTVTETAKNGOKYTLHFDR---LGRMPTW-- 386
Db 4237 SKL-----NKNYSDIPQVPRP--ISTPSIPDSRNL--DRNSPEGSAIPEHE 4285
Qy 387 --VISPSKKGYIE-----OYGT-----402
Db 4286 FSTFNPESVGHRRKAVACSVAPNLPPLPPSPDSISQKPSWDEFDYDTKVVDLDPL 4345
Qy 403 --DPVTGKA-PYSATSVLKL-----GYLWIDIEDFTPRV-----AHSPSF 440
Db 4346 SKKPLEKPSQPYSAESELSEVQSLSSFQSCDNDGHWDTSDMWPVPLDQIEFPNY 4405
Qy 441 DHLIGTTLREDAPIALKTPH 460
Db 4406 E-----VIDEQTPLYSADPN 4420

RESULT 13
COA2_PAVPN STANDARD; PRT; 749 AA.
AC P22964; 1991 (Rel. 19, Created)
DT 01-AUG-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE COAT PROTEIN VP1 [CONTAINS: COAT PROTEIN VP2] (VERSION 2).
OS Porcine parvovirus (strain NADL-2) (ppv).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90010964; Pubmed=2794971;
RA Ranz A.I., Mancius J.J., Diaz-Aroca E., Casal J.I.;
RT "Porcine parvovirus: DNA sequence and genome organization.";
RL J. Gen. Virol. 70:2541-2553(1989).
CC 1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC 2- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC 3- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00623; BAA00502.1;
DR PIR; B33302; VCPVPP.
DR HSSP; P30129; 4DPV.
DR INTERPRO; IPR001403;
DR PFAM; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein. COAT PROTEIN VP1.
FT CHAIN 1 749

```

```

FT CHAIN 171 749 COAT PROTEIN VP2.
FT DOMAIN 195 210 POLY-GLY.
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 749 AA; 82872 MW; 6CA13CB97BE93418 CRC64;

Query Match 4.2%; Score 105; DB 1; Length 749;
Best Local Similarity 19.9%; Pred. No. 4.5; Indels 222; Gaps 30;
Matches 116; Conservative 56; Mismatches 189;

Qy 20 SPVTSEYTSVREAPFGYKPSKES-----IENLKDVKVENIVLILLENFSFNILG 69
Db 120 SPTTSQOEVRRSRKH-FGSKPPGKRPAPRHIFINLAKKKAG-----TSNTNSMSME 173
Qy 70 GYRROGLDNPIN-----NGPFCN-----87
Db 174 NVEQH---NPINAATELSATNGSGGGGGGGGAGGVGVSTGFSFNQTEFYVLGGLV 230
Qy 88 -----YK-----NASDSSCKYC-----TQ-----AKOYDSV 109
Db 231 RITAHASRLHNLNPEHETKYRHLVNSGSGAGVQVDDAHTQMTWPSLIDANAGVW 290
Qy 110 FNDPDHSTVGNLEFYG-----TYTPNNGAIASGKGVADOS 145
Db 291 FNPADQQLISNNTEINLVSEFQAIINVLKLTITESATSPPTKIYNNDLTASLHVALDTN 350
Qy 146 GFNLQNLNDYKPLAPEATQVNGYTYEEVPTLVLDVDEFTTENSFSCVPGPTNPRL 205
Db 351 NTL-----PYTPAAPRSET---LGFY---PWLP-----KPTQYRYLSCIRNLNP-- 391
Qy 206 CALAGTAAGHGKNDDEFLNYGSSKSIPEANKEGYSWLNVDG---TNG---EFEPDSLFF 260
Db 392 -----TYTGQSQQTDSIQTGLHSDIMFYTIENAVPIHLLRTGDEFTGIYHFDTKPLK 446
Qy 261 TYVQNTSRNVVVENFQ-----DAYLGLVLPKFSINPSCCGTNT-NSMHPTGNVSYGE 314
Db 447 THSQVNRSLGLUPPALLTEPTTEGQDHGCTLP-----AANTRGYHOTINNSYTE 496
Qy 315 VFVKQIYDAIRGP-QMDKTLFLFYDETCGFYDHVPPPLAVRPNLTVTETAKNGQ-KY 372
Db 497 A-----TAIRPAQVGYNTPTMNFYSNGGPPPLTPIVPTA-----DTQYNDDEPGAIRF 545
Qy 373 TL-----HFDRLGGRMPTWVISPSKKGYI---EYVG-----TDPVTGK 408
Db 546 TMDYQHGHTTSQELERYTFNQSKCGRAPKQENQQAPLNLENTNGTLLPSDPIGK 605
Qy 409 P-----APYSA-----TSVLKYLGLWIDIE---DFTPRV 434
Db 606 SNMHEMTLNTYGLPLTALNNTPAVPFNGQIWDKELDTDLKPLR 648

RESULT 14
BGAL_LACSK STANDARD; PRT; 625 AA.
AC Q46846; 1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-GALACTOSIDASE LARGE SUBUNIT (EC 3.2.1.23) (LACTASE).
GN LACL.
OS Lactobacillus sake.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
RN [1]
RP SEQUENCE FROM N.A.

```





Db 477 A-----TAIRPAQVGYNTPYMNFYSNGGPFELTPIVPTA-----DTQYNDEPNGAIRF 525  
Qy 373 TL-----HFDRLGGRMPTWVISPYSKGYI--BOYG-----TDPVTGK 408  
Db 526 TMDYQHGLTTSSOELERTFNPQSKCGRAPKQOFNOQAPLNLNTNNGTLLPSPDPGCK 585  
Qy 409 P-----APYSA-----TSVLKTLGYLWDIE---DETDRV 434  
Db 586 SNMHFWMTLNTYGPLTALANTAPVFPFNGQIWDKELDTDLKPRL 628

Search completed: February 26, 2001, 19:00:16  
Job time: 17410 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2001, 14:04:08 ; Search time 64.74 Seconds  
(without alignments)  
245.071 Million cell updates/sec

Title: US-09-426-072-2

Perfect score: 2501

Sequence: 1 MKSTALLTGGLASIGLAS.....GTTLRDAPIALKTPHPSV 464

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A. Geneseq\_36.\*  
1: /cgn2\_2/cgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /cgn2\_2/cgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /cgn2\_2/cgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /cgn2\_2/cgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /cgn2\_2/cgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /cgn2\_2/cgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /cgn2\_2/cgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /cgn2\_2/cgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /cgn2\_2/cgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /cgn2\_2/cgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /cgn2\_2/cgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /cgn2\_2/cgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /cgn2\_2/cgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /cgn2\_2/cgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /cgn2\_2/cgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /cgn2\_2/cgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /cgn2\_2/cgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /cgn2\_2/cgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /cgn2\_2/cgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /cgn2\_2/cgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /cgn2\_2/cgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126.5	5.1	483	11 R07445	Secretory signal p
2	111.5	4.5	1041	18 W11866	ATH1 gene product,
3	109.5	4.4	579	14 R38697	PSY875 swine parvo
4	109.5	4.4	1041	20 V16788	S. cerevisiae acid
5	107	4.3	532	21 V58162	Adeno associated v
6	107	4.3	586	21 V58161	Adeno associated v
7	107	4.3	724	21 V58160	Adeno associated v
8	107	4.3	1257	13 R29028	Bacillus thuringie
9	107	4.3	1257	13 R28811	BT toxin 33F2. Ba
10	107	4.3	1257	13 R29518	BT toxin 33F2. Ba
11	107	4.3	1257	14 R44203	Bacillus thuringie
12	107	4.3	1257	18 W13886	33F2 toxin. Bacil

13	106	4.2	1257	13 R20068	B. thuringiensis to
14	105.5	4.2	1041	20 Y29830	Saccharomyces cere
15	105.5	4.2	1122	16 R64927	Cyathesin protein
16	104.5	4.2	737	16 R70186	Arg-gingipain-1.
17	104.5	4.2	737	18 W34846	Arg-gingipain high
18	104.5	4.2	737	21 Y67395	Arg-gingipain-2 am
19	104.5	4.2	991	16 R77313	Porphyromonas ging
20	104.5	4.2	1687	17 R96033	P. gingivalis haem
21	104.5	4.2	1687	19 W69495	Haemagglutinin pro
22	104.5	4.2	1704	16 R70188	Arg-gingipain-2 pr
23	104.5	4.2	1704	18 W34843	Arg-gingipain high
24	104.5	4.2	1704	21 Y67396	Arg-gingipain-2 am
25	104	4.2	783	21 Y75533	Neisseria meningit
26	104	4.2	783	21 Y75534	Neisseria meningit
27	103.5	4.1	579	14 R38702	Swine parvovirus B
28	103.5	4.1	579	17 R99721	Swine parvovirus B
29	103.5	4.1	598	8 P70500	Pig parvo virus B
30	103.5	4.1	1290	18 W07609	Rat von Ebner's gl
31	103	4.1	636	20 W89548	Bacillus sp. alkal
32	103	4.1	686	5 P40068	Sequence of a porc
33	102	4.1	433	13 R26274	Alkali-protease ya
34	102	4.1	433	19 W61495	Modified Bacillus
35	102	4.1	433	20 W95698	Bacillus sp. Lion
36	102	4.1	433	21 Y69207	Amino acid sequenc
37	102	4.1	433	21 Y44619	Bacillus Lion Y en
38	101.5	4.1	362	21 Y44497	Bacillus subtilisi
39	101.5	4.1	746	17 R93270	GSP-INL fusion int
40	101.5	4.1	1706	18 W24786	Prtr antigenic pro
41	101	4.0	2893	19 W98828	H. pylori GHPO 148
42	101	4.0	2893	19 W71556	Helicobacter polyp
43	100	4.0	1213	18 W57535	H. pylori ORF 07ee
44	96.5	3.9	410	21 Y32309	Soybean acid triac
45	96.5	3.9	493	20 W83125	PrTIIR45 Arginine

#### ALIGNMENTS

RESULT	1
R07445	ID R07445 standard; protein; 483 AA.
XX	AC R07445;
XX	DT 01-FEB-1991 (first entry)
XX	DE Secretory signal peptide and protein with repressible acid
XX	DE phosphatase activity.
XX	KW Acid phosphatase; yeast.
XX	OS Saccharomyces cerevisiae.
XX	XX Key Location/Qualifiers
XX	FT Peptide 1..21
XX	FT Protein /label=Secretory signal peptide
XX	FT 22..483
XX	FT /label=Protein has repressible acid phosphatase activity
XX	PN JP02234681-A.
XX	PD 17-SEP-1990.
XX	PF 09-MAR-1989; 89JP-0057234.
XX	PR 09-MAR-1989; 89JP-0057234.
XX	PA (IKEM ) IKEDA MOHANDO KK.
XX	DR WPI; 1990-325618/43.
XX	DR N-PSDB; Q06266.
XX	PT DNA contg. base sequences with promoter activity - used for

PT expression and secretory prodn. of pharmacologically important  
 XX protein gene  
 PS Claim 1; Page 553; 11pp; Japanese.  
 CC Protein has repressible acid phosphatase activity, and carries a  
 CC secretory signal region allowing it to be produced in high yield into  
 CC the extracellular environment.  
 XX Sequence 483 AA;  
 SQ

Query Match 5.1%; Score 126.5; DB 11; Length 483;  
 Best Local Similarity 23.0%; Pred. No. 0.0054;  
 Matches 67; Conservative 47; Mismatches 112; Indels 65; Gaps 14;

QY 134 ATASKKVADQ-SGFNLQNLN-----DYPKLAPEEATROVMGYTEEVPTLVLDL 182  
 DB 48 aaasavtahrtdvkgavnrffvflentdydkavede-----slawl 92

QY 183 VDEFTTNSWFSVCPGPTNPRLCALAGTAAGHKNDDDFLNYGSSKSIPEAANEKGV 242  
 DB 93 akqgisitnyws-lthpsepylasvagy--falddrflsmnsnvnivlditknis 149

QY 243 W-----LNYDGTNG-----EPEPDSLFYTVNQTsr-SNVVPVENFQD 280  
 DB 150 waeyqehipytygfqfynsqetfandyrkhnplilfdnvisdkarlaniksfedntd 209

QY 281 AYLGVLPKFSYNIPSCCGTNTNSMHTPGNVSYGEVFKQIYDAIROGPQWDR-TLLFTY 339  
 DB 210 lknktlpqyafitpnm-----tndghds-nikvagdwsksflqlpsddyfmkdtviltf 264

QY 340 D--ETGQFVDHVPPL--AVRPDLTYTETAKNGOKYTLFDLGRMPTW 386  
 DB 265 denetgyiknkvfsilgvgvipdhkgt-----sddtfydhysqlatveanw 311

RESULT 2  
 W11866  
 ID W11866 standard; Protein; 1041 AA.  
 AC W11866;  
 XX  
 DT 18-APR-1997 (first entry)  
 DE ATH1 gene product, Athlp.  
 XX  
 KW Yeast ATH1; Athlp; vacuolar acid trehalase; hydrolysis; trehalose;  
 KW baking; frozen dough; dehydrated yeast; brewing; ethanol fuel;  
 KW molasses; corn syrup; freezing; wine; fermentation.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN US5587290-A.  
 XX  
 PD 24-DEC-1996.  
 XX  
 PF 26-JUN-1995; 95US-0494714.  
 XX  
 PR 26-JUN-1995; 95US-0494714.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Destruelle M, Holzer H, Klionsky D;  
 XX WPI; 1997-064793/06.  
 DR N-PSDB; T61372.  
 XX  
 XX Mutant yeast deficient in prodn. of vacuolar acid trehalase - has  
 PT improved stress tolerance, esp. for dehydration, freezing and  
 PT alcohol concn., useful in baking, fermentation, etc.  
 XX  
 PS Disclosure; Column 17-24; 17pp; English.

XX This sequence is encoded by the yeast ATH1 coding sequence and represents  
 CC Athlp which is a vacuolar acid trehalase. Yeast mutants which are  
 CC deficient in the production of functional Athlp have reduced ability  
 CC to hydrolyse trehalose. These mutants are useful in baking (partic. in  
 CC frozen dough and dehydrated yeast products), in brewing, as a source of  
 CC trehalose (used as a protectant in foods and pharmaceuticals) and for  
 CC production of ethanol fuel from molasses or corn syrup. The yeast  
 CC mutants have improved tolerance of dehydration and freezing. They can  
 CC tolerate higher levels of ethanol and can grow at high cell densities  
 CC over a range of fermentable sugar concentrations, and so can produce  
 CC drier wines and complete fermentation more quickly.  
 XX Sequence 1041 AA;  
 SQ

Query Match 4.5%; Score 111.5; DB 18; Length 1041;  
 Best Local Similarity 19.7%; Pred. No. 0.4;  
 Matches 89; Conservative 64; Mismatches 140; Indels 159; Gaps 24;

QY 22 VTSEY---TSVREAPFGYKPGSKESIEIENLKRVENIVMLILENRS-----FDNILGG 70  
 DB 446 vdyeihinvdvamasfsiylngheidd--eylryttwpliknaaaffaykvkynsslgl 503

QY 71 VRRQGLDNP-----INNGPFCNYKNASDPSSGKYCTQAKDYSDVFNPDHSHVTGNLEF 124  
 DB 504 yetynltqdefanhnngaf---tnagiktilkwatdighnhgevvdpkswseiskdi-- 558

QY 125 YGYTNNCAIASGKVADOSGFLNAQLNDYPKLAPEEATROV--MGYYTEEEVPTLVLDL 182  
 DB 559 ---yipr-----sssnitleysg-----msssvveikqadvtlmvyplgyindesi--lnna 604

QY 183 VDEFTTNSWFSV--PGPTNPRLCALAGTAAGHG-----K 217  
 DB 605 ikdlyyysergsasgampatyvfvaagaag-llnhgssqsylykvslylrapfaqfseq 663

QY 218 NDDDFLNYGSSKSI-FAEANEK-----GVSWLNVDGTNGEPEPDSLFYTVNQTsrS 269  
 DB 564 sddnflngltqpaifltangflqslfllgtirly---syevdpdt---kkinrlirf 717

QY 270 NVVPVENFQDAYLGVLP-----KFSYINPSCCGTNTNSMHTPGNVSYGEVFKQIYDA 323  
 DB 718 n--pie-----lpllpaggialrnfkymnp----- 739

QY 324 IROGPQWDRKTLLETTYDETGGFYDH-----VPPPLAVRPDLNLTYYTAKNGOKYT 373  
 DB 740 -----vldliidhngitvhksgdvpihiklpnrslhdqddinfyngsenetkpn 789

QY 374 LH---FDRLLGRMPTWVISPYSKKGYIEQYGT 402  
 DB 790 lerrdvdrvgdpm-----rmdrygt 809

RESULT 3  
 R38697  
 ID R38697 standard; Protein; 579 AA.  
 AC R38697;  
 XX  
 DT 25-NOV-1993 (first entry)  
 DE PSY875 swine parvovirus B gene product.  
 XX  
 KW Attenuated; vaccine; herpes virus; non-primate; liver; safer; IBR;  
 KW infectious bovine rhinotracheitis; MDV; Marek's disease virus; fowl;  
 KW pseudo-rabies; swine.  
 XX  
 OS Swine parvovirus.  
 XX  
 PN US523424-A.  
 XX  
 PD 29-JUN-1993.  
 XX



Sequence 532 AA;

Query Match 4.3%; Score 107; DB 21; Length 532;  
Best Local Similarity 19.1%; Pred. No. 0.36;  
Matches 66; Conservative 35; Mismatches 113; Indels 13

QY 118 TGNNLE-----FYGTYTPNNGAIASGKVVADQSGFLNAQLNDYPKLAPEEATROVM 168

Db 206 tgnnfteytnfeevpfhssfapqnlfkplanplvdqylrfvstnn--tqqvqfnknl 262

QY 169 GYYTEEEVPTLVLDLVEFTTFNSWFSCVPGPTPNRLCALAGTAAGHGKNDDEFLNYGIS 228

```

Db 263 grya-----ntyknwf---pypmgnrtq-----gwnlqsgvnr-----a 292
QY 229 SKSIFEAAKEGYSMLNDGNGEFPEDSLFTTYVNOTSRNVVPVENFF-----QDAYLG 284
Db 293 svsafattnr-----melegasvqppqpngnt--nnlqgsntya lentmifnsgpanpg 345

```

QY 285 VLPKFSYINPSCCGTNTNSMHP TGNVSY-----GEV 315

Db 346 tta--tylegnmlitsetqpvrwaynvvggqmatnnqssstapatqtynlqei vpsv 403

QY 316 FVKQIYDAIRQGPQWDKTLFITYDETGGFY-----DHVPPPLAVR-----PDNI. 360

```

Db      404 wmer--dvylqqpiwak-----ipetqahfhpspamgqfalhhpppmm}ikntpyvan; 455

```

QY 361 TYTETAKNGOKYTLHFDRLGGRMPTWVISPYSKKGYTEOYGTDPVT 406

Db 456 t-----sfdsdvnvssfi tnvst gvv 176 .

RESULT	6
Y58161	

LD Y58161 standard; Protein; 588 AA.  
XX

AC Y58161;  
XX XX

XX  
000000 (111111 entry)  
XX

PI Chlorini JA, Kotin RM;

XX WPI: 2000-062707/05.  
 XX N-PSDB: Z49213.  
 XX Adeno-associated virus 5 based vectors and particles, useful for gene  
 PT therapy -  
 XX Claim 25; Page 79-80; 91pp; English.  
 XX This sequence represents the VP2 capsid protein of  
 CC adeno associated virus type 5 (AAV5). The invention relates  
 CC to vectors comprising a pair of AAV5 inverted terminal repeats  
 CC (ITRs) with a promoter between the ITRs. The vector may comprise  
 CC the viral genome, or subregions thereof, including sequences  
 CC encoding Rep proteins and capsid proteins, and is encapsidated  
 CC in an AAV5 particle. The non-structural Rep proteins Rep40  
 CC (Y58613), Rep52 (Y58168), Rep8 (Y58164) and Rep78 (Y58159) are  
 CC involved in regulation of replication and transcription, in addition to  
 CC the production of progeny genomes. Rep68 and Rep78 are also associated  
 CC with the stable integration of the viral genome into human chromosomes.  
 CC The three types of capsid protein VP1 (Y58160), VP2 (Y58161) and VP3  
 CC (Y58162) assemble to form an icosahedral capsid, and differ from each  
 CC other by the use of alternative splicing and an unusual translation  
 CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2  
 CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are  
 CC likely to utilise distinct cellular receptors and are serologically  
 CC distinct. In a gene therapy application, therefore, AAV5 would allow for  
 CC transduction of a patient who already possess neutralising antibodies  
 CC either as a result of natural immunological defence or from prior  
 CC exposure to AAV2 vectors. The vectors may be useful for transducing  
 CC erythroid progenitor cells or cells lacking heparin sulphate  
 CC proteoglycans, which is very inefficient with AAV2-based vectors.  
 CC The vectors may also be useful for transducing cells with a nucleic  
 CC acid of interest in order to produce cell lines that could be used to  
 CC screen for agents that interact with the gene product of the nucleic  
 CC acid of interest. In addition to transduction of other cell types,  
 CC transduction of erythroid cells would be useful or the treatment of  
 CC cancer and genetic diseases which can be corrected by bone marrow  
 CC transplants using matched donors.  
 XX Sequence 588 AA;  
 SQ

Query Match 4.3%; Score 107; DB 21; Length 588;  
 Best Local Similarity 19.1%; Pred. No. 0.42;  
 Matches 66; Conservative 35; Mismatches 113; Indels 132; Gaps 16;

Oy 118 TGNLLE-----FVGTTPNGAIGASGKVVADQSGFLNAQLNDPKLAPEATRQVM 168  
 Db 262 tgmtefyntfeevpfhssfapsqnlflklaplvdqlyrfvstnn---tgvgqfknla 318  
 Oy 169 GYYTEEVPTLVLDVETFTFNSFSCVPGPTPNRICALAGTAGCHKNDDELFLYNGIS 228  
 Db 319 gyya-----ntyknwf---pgpmgrtg-----gwnlgsgvnr-----a 348  
 Oy 229 SKSIFEANERKGSWLNVDGTNGEPEFPOSLEFTYVNTSRSNVVPVENFF-----QDAYLG 284  
 Db 349 svsafattnr-----meleqasyvppqpngmt---nnlgsntyalentmifnsqpnpq 401  
 Oy 285 VLPKFSYNPCSCGTNTNSMHPGTGNVSY-----nnlgsntyalentmifnsqpnpq 315  
 Db 402 tta--tylegmmlitsetqpnrvaynvggmatnngssstapatytnlqlvpgsv 459  
 Oy 316 FVKQIYDAIRGQPOWDXTLITFDYDTGTFY-----DHWPPPLAVR-----PDNL 360  
 Db 460 wmer--dylqgplwak-----ipetgahfhpspanggfglkhpmmllikntpvggni 511  
 Oy 361 TYPETAKNQXYTLHFDRLGRRMPTWVISPYSKKGYTEQYIGTDPVT 406  
 Db 512 t-----sfsdvpvssfitgystgqvt 532

RESULT 7

Y58160  
 ID Y58160 standard; Protein; 724 AA.  
 XX AC Y58160;  
 XX DT 07-MAR-2000 (first entry)  
 XX DE Adeno associated virus AAV5 capsid protein VP1.  
 XX KW Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;  
 KW promoter; Rep protein; capsid protein; regulation; transcription;  
 KW replication; chromosomal integration; tissue tropism; cellular receptor;  
 KW gene therapy; neutralising antibody; erythroid progenitor cell;  
 KW transduction; cancer; genetic disease; VP1.  
 XX OS Adeno associated virus type 5.  
 XX PN WO9961601-A2.  
 XX PD 02-DEC-1999.  
 XX PF 28-MAY-1999; 99WO-US11958.  
 XX PR 28-MAY-1998; 98US-0087029.  
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
 XX PI Chlorini JA, Kotin RM;  
 XX WPI: 2000-062707/05.  
 XX DR N-PSDB: Z49212.  
 XX PT Adeno-associated virus 5 based vectors and particles, useful for gene  
 therapy -  
 XX Claim 23; Fig 5; 91pp; English.  
 XX This sequence represents the VP1 capsid protein of  
 CC adeno associated virus type 5 (AAV5). The invention relates  
 CC to vectors comprising a pair of AAV5 inverted terminal repeats  
 CC (ITRs) with a promoter between the ITRs. The vector may comprise  
 CC the viral genome, or subregions thereof, including sequences  
 CC encoding Rep proteins and capsid proteins, and is encapsidated  
 CC in an AAV5 particle. The non-structural Rep proteins Rep40  
 CC (Y58613), Rep52 (Y58168), Rep8 (Y58164) and Rep78 (Y58159) are  
 CC involved in regulation of replication and transcription, in addition to  
 CC the production of progeny genomes. Rep68 and Rep78 are also associated  
 CC with the stable integration of the viral genome into human chromosomes.  
 CC The three types of capsid protein VP1 (Y58160), VP2 (Y58161) and VP3  
 CC (Y58162) assemble to form an icosahedral capsid, and differ from each  
 CC other by the use of alternative splicing and an unusual translation  
 CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2  
 CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are  
 CC likely to utilise distinct cellular receptors and are serologically  
 CC distinct. In a gene therapy application, therefore, AAV5 would allow for  
 CC transduction of a patient who already possess neutralising antibodies  
 CC either as a result of natural immunological defence or from prior  
 CC exposure to AAV2 vectors. The vectors may be useful for transducing  
 CC erythroid progenitor cells or cells lacking heparin sulphate  
 CC proteoglycans, which is very inefficient with AAV2-based vectors.  
 CC The vectors may also be useful for transducing cells with a nucleic  
 CC acid of interest in order to produce cell lines that could be used to  
 CC screen for agents that interact with the gene product of the nucleic  
 CC acid of interest. In addition to transduction of other cell types,  
 CC transduction of erythroid cells would be useful or the treatment of  
 CC cancer and genetic diseases which can be corrected by bone marrow  
 CC transplants using matched donors.  
 XX Sequence 724 AA;  
 SQ

Query Match 4.3%; Score 107; DB 21; Length 724;  
 Best Local Similarity 19.1%; Pred. No. 0.58;

Matches	66; Conservative	35; Mismatches	113; Indels	132; Gaps	16;	Matches	97; Conservative	73; Mismatches	166; Indels	166; Gaps	23;
QY 118	TCNNLE-----FYCTYTPNNGAIAASGVYVADQSGFLNAQINDYPKLAPEEATROVM	168				QY 6	LLTCLGLLASLG-LASPYTSEYTSVRADPGYKPGSKESIEINKDKVENIVWLLEN---	61			
Db 398	tgnnfeytfeevfhssafsqnlkianplvdqyllyrfvstnn---tgvgqfnkula	454				Db 74	visviltipevctvasaastvfiwpkifgdkpnakniifeelkpgqlealqidditnyqd	133			
QY 169	GXYTEEEVPTLDVDEFTTNSWFSVCPTNPNRLCALAGTAAGHKNDLDFLNYGIS	228				QY 62	-----RSFDNILGGVRRQGLDNPINNGPFCNYKNASDPSSGKYCTQAKDYDVSFEND----	112			
Db 455	guya-----ntyknwf---ppgmgrtq-----gnlsgvnr-----a	484				Db 134	ainqkfkfslqktinlytv--aidnodyvtak-----tqlenlnsllsdisif	180			
QY 229	SKSIFEAAEKGVSNLYDGTNGEFPDLSFTYVYNQTSRNVVPVENFF-----QDAYL	284				QY 113	PDHSVTGNLLEFYCTYTPNNGAIAAGKVY-ADQSGFLNAQINDYPKLAPEEATROVMGY	170			
Db 485	svsafatnr-----melegasyqvpqpngmt--nnlqgsntyalentmifnsqpanpg	537				Db 181	ipegyetg-glpvyamvanahillirdaivnaeklgfsdkevthkky-----ikmtihn	234			
QY 285	VLPKFSVINPSCCGTNTSMHPTGVSY-----	315				QY 171	YTEEEVPTLDVDEFTTF--NSWFSVCVPGPTNPNRLCALAGTAAGHKNDLDFLNYGIS	228			
Db 538	tta--tylegmllisesetqpvnrvaaynvvgqmatnqsgstapagtynlqetvpsv	595				Db 235	hteavikafingldkfksldvnsy-----nkkany---	264			
QY 316	FKVQIYDAIRQGPQWDKTLFTYDETGFGY-----DHVPPPLAVR-----PDNL	360				QY 229	SKSIFEAAEKGVSNLYDGTNGEFPDLSFTYVYN-----QTSRNVVPVENF	277			
Db 596	wmer--dvylqgpiwak-----ipetgahfhpspamggfklkhpmmalikntpvpgni	647				Db 265	ikgmtemvldivalwptfdpdykveleletrtisspiyqvpvknmqntssivpsdlf	324			
QY 361	TYTETAKNGQKVTLHFDRLGRMPTWVISPYSKGYLEOXTDPTV	406				QY 278	FQDAYLGVLPK--FS-----YINPSCCGT-----YINPSMHPTGN	309			
Db 648	t-----sfadvpvsfitystqgvt	668				Db 325	h---yqgdlvklefstrtdndagiakiftgirtfykspnthelyhvdfsyntqs---sgn	378			
RESULT 8						QY 310	VSYG-----EVEVKQIYD-----AIRQGPQWDKTLF	336			
R29028						Db 379	isrgsenpildlnpiistcirmsfykaigssvlnfkdgtygafagap-----	430			
XX R29028						QY 337	ITYDETGFGYDH--VPPPLAVRPNLTYYETAKNG-----OKYTLHFDRLGRMPTWVI	388			
AC R29028						Db 431	-----tgadwhsfiesdgapeghklnlytspgdtrldfinvlytlistptinelsteki	485			
DT 20-APR-1993 (first entry)						QY 389	SPY-SKKGYIEQYGTDPVTGKP	409			
DE Bacillus thuringiensis toxin protein 33F2.						Db 486	kgfpaekgyiknggimkygkp	507			
XX Toxin protein; ant.						RESULT 9					
XX Bacillus thuringiensis strain PS33F2.						R28811					
XX W09220802-A.						ID R28811	standard; Protein; 1257 AA.				
XX 26-NOV-1992.						XX R28811;					
XX 22-MAY-1992; 92WO-US04316.						DT 30-MAR-1993 (first entry)					
XX 22-MAY-1991; 91US-0703977.						DE BT toxin 33F2.					
PR 25-NOV-1991; 91US-0797645.						XX nematode worms; nematocides; nematocidal toxin; agriculture; plants;					
PR 12-MAY-1992; 92EP-0304228.						XX crops; pests; CryV proteins.					
XX (MYCO ) MYCOGEN CORP.						XX Bacillus Thuringiensis.					
PI Kennedy MK, Meier H, Payne JM, Randall JB, Ulick HJ;						XX W09219739-A.					
XX WPI; 1992-415780/50.						XX 12-NOV-1992.					
DR N-PSDB; Q31411.						XX 01-MAY-1992; 92WO-US03624.					
XX Toxin proteins isolated from Bacillus thuringiensis - for controlling						XX 03-MAY-1991; 91US-0693018.					
PT ants. e.g. fire, carpenter, Argentine and pharaoh ants						PR 31-JAN-1992; 92US-0830050.					
XX Disclosure; Page 45; 71pp; English.						PR 23-APR-1992; 92US-0871510.					
XX Bacillus thuringiensis toxin protein 33F2 is useful as a method of						XX (MYCO ) MYCOGEN CORP.					
CC biological control of ants, e.g. fire ants, carpenter ants,						XX Focerrada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;					
CC Argentine ants and pharaoh ants, as an alternative to chemical						XX WPI; 1992-398866/48.					
CC insecticides.						DR N-PSDB; Q30937.					
XX Sequence 1257 AA;						XX New genes and toxins against nematodes - obt'd. from Bacillus					
SQ						PT Thuringiensis isolates with nematocidal activity					

Query Match 4.3%; Score 107; DB 13; Length 1257;  
 Best Local Similarity 19.3%; Pred. No. 1.4;



XX PS Claim 1(b) : Page 54; 77pp; English.  
 XX CC This sequence represents the *Bacillus thuringiensis* delta-endotoxin  
 CC 33F2 having nematocidal activity. It was decoded from the appropriate  
 CC DNA. Intact cells expressing this toxin may be used in nematocidal  
 CC compns, either sprayed on plants to kill agricultural crop pests,  
 CC or taken as a drug to kill gastro-intestinal worms in mammalian hosts.  
 XX SQ Sequence 1257 AA;

Query Match 4.3%; Score 107; DB 13; Length 1257;  
 Best Local Similarity 19.3%; Pred. No. 1.4; Indels 166; Gaps 23;  
 Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;  
 QY 6 LITGLGLLASLG-LASPVTSYTSVREAPFGYKPGSKESINLKDKNVENIVWLILEN--- 61  
 Db 74 vlsvltlilpevgtsaastivsfwfkfgdkpnaknifeelkpgleallqgditnyqd 133  
 QY 62 ---RSPDNILGGVRRQGLDNPINNGPFCNYKNASDPSSGKYCTQAKDYDSVFN--- 112  
 Db 134 aingkkfidsiqklnlytv--aidndyvtak-----tqlenlsiltsdisif 180  
 QY 113 -PDHSVTGNLLEFYGYTTPNNGAIASGVV-ADQSGFLNAQLNDYPKLAPEEATROVMGY 170  
 Db 181 ipegvetyg-glpyyavnanahillirdaivnaeklgfsdkevthkky-----ikmthn 234  
 QY 171 YTEEEVPTLVLDVDEFTTF--NSWFSCVPGPTNPRLCALAGTAAGHGKNDDDLNYGIS 228  
 Db 235 hceavikafingldkfksldvnsy-----nkkany--- 264  
 QY 229 SKSIFEAAANEKGVSWLNYDGTNGEPEPDSLFFTYVN-----QTSRSNVVPVNEF 277  
 Db 265 ikgmtemvldlvalwptfdpdykveleftrtisspiyqvpknmqntssivpsdlf 324  
 QY 278 FQDAVLGVLPK--FS-----YINPSCCGT-----NTNSMHPGTGN 309  
 Db 325 h---yggdlvklefstrtdnglakifgtrntfykspnthethyhdvfsyntqs---sgn 378  
 QY 310 VSYG-----EVFVKQIYD-----AIROGPQWDKTLF 336  
 Db 379 isrgssnplidlnplistcinsfykagagssvlnfkdgctgyafagap----- 430  
 QY 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKTYLHFDRLGGRMPTWVI 388  
 Db 431 -----tggawdhsfiesdgapeghklnlytspgdtlrdfinvytlstptinelsteki 485  
 QY 389 SPY-SKKGIEQYGTDPVTGKP 409  
 Db 486 kgfpaekgyiknqgmkygk 507

RESULT 10  
 R29518  
 ID R29518 standard; Protein; 1257 AA.  
 XX AC R29518;  
 XX DT 20-APR-1993 (first entry)  
 XX DE BT toxin 33F2.  
 XX KW nematode worms; nematocidal; nematocidal toxin; agriculture; plants;  
 XX KW crops; pests; CryV proteins.  
 XX OS *Bacillus thuringiensis*.  
 XX PN EP517367-A.  
 XX PD 09-DEC-1992.  
 XX XX 01-MAY-1992; 92EP-0303969.

XX PR 03-MAY-1991; 91US-0693018.  
 PR 31-JAN-1992; 92US-0830050.  
 PR 23-APR-1992; 92US-0871510.  
 XX PA (MYCO ) MYCOGEN CORP.  
 XX PI Foncerrada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;  
 XX DR WPI; 1992-408829/50.  
 DR N-PSDB; Q32161.  
 XX PT Nematocidal toxins from *Bacillus thuringiensis* - useful for  
 PT control of animal or plant parasites, deoxyribonucleic acid  
 PT coding sequences, transformed hosts and transgenic plants  
 XX PS Claim 1(b); Page 31; 57pp; English.  
 XX CC This sequence represents the *Bacillus thuringiensis* delta-endotoxin  
 CC 33F2 having nematocidal activity. It was decoded from the appropriate  
 CC DNA. Intact cells expressing this toxin may be used in nematocidal  
 CC compns, either sprayed on plants to kill agricultural crop pests,  
 CC or taken as a drug to kill gastro-intestinal worms in mammalian hosts.  
 XX SQ Sequence 1257 AA;

Query Match 4.3%; Score 107; DB 13; Length 1257;  
 Best Local Similarity 19.3%; Pred. No. 1.4; Indels 166; Gaps 23;  
 Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;  
 QY 6 LITGLGLLASLG-LASPVTSYTSVREAPFGYKPGSKESINLKDKNVENIVWLILEN--- 61  
 Db 74 vlsvltlilpevgtsaastivsfwfkfgdkpnaknifeelkpgleallqgditnyqd 133  
 QY 62 ---RSPDNILGGVRRQGLDNPINNGPFCNYKNASDPSSGKYCTQAKDYDSVFN--- 112  
 Db 134 aingkkfidsiqklnlytv--aidndyvtak-----tqlenlsiltsdisif 180  
 QY 113 -PDHSVTGNLLEFYGYTTPNNGAIASGVV-ADQSGFLNAQLNDYPKLAPEEATROVMGY 170  
 Db 181 ipegvetyg-glpyyavnanahillirdaivnaeklgfsdkevthkky-----ikmthn 234  
 QY 171 YTEEEVPTLVLDVDEFTTF--NSWFSCVPGPTNPRLCALAGTAAGHGKNDDDLNYGIS 228  
 Db 235 hceavikafingldkfksldvnsy-----nkkany--- 264  
 QY 229 SKSIFEAAANEKGVSWLNYDGTNGEPEPDSLFFTYVN-----QTSRSNVVPVNEF 277  
 Db 265 ikgmtemvldlvalwptfdpdykveleftrtisspiyqvpknmqntssivpsdlf 324  
 QY 278 FQDAVLGVLPK--FS-----YINPSCCGT-----NTNSMHPGTGN 309  
 Db 325 h---yggdlvklefstrtdnglakifgtrntfykspnthethyhdvfsyntqs---sgn 378  
 QY 310 VSYG-----EVFVKQIYD-----AIROGPQWDKTLF 336  
 Db 379 isrgssnplidlnplistcinsfykagagssvlnfkdgctgyafagap----- 430  
 QY 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKTYLHFDRLGGRMPTWVI 388  
 Db 431 -----tggawdhsfiesdgapeghklnlytspgdtlrdfinvytlstptinelsteki 485  
 QY 389 SPY-SKKGIEQYGTDPVTGKP 409  
 Db 486 kgfpaekgyiknqgmkygk 507

RESULT 11  
 R44203  
 ID R44203 standard; Protein; 1257 AA.  
 XX XX  
 AC R44203;

XX 09-DEC-1995 (first entry)  
 XX Bacillus thuringiensis crystal protein 33f2.  
 DE Crystal protein; delta endotoxin; acaricide; pesticide;  
 XX biological control agent; transgenic plant; crop improvement.  
 KW Bacillus thuringiensis (isolate 33f2).  
 XX US5262158-A.  
 XX 16-NOV-1993.  
 XX 30-APR-1991; 91US-0693210.  
 XX 30-APR-1992; 92US-0876280.  
 PR 30-APR-1991; 91US-0693210.  
 PR 13-SEP-1991; 91US-0759248.  
 PR 30-SEP-1991; 91US-0768141.  
 XX (MYCO ) MYCOGEN CORP.  
 PA Bagley AL, Cannon RJC, Payne JM;  
 PI WPI: 1993-377387/47.  
 XX N-PSDB; Q51689.  
 DR Controlling acaride pests e.g. two spotted spider mite - utilising  
 PT Bacillus thuringiensis delta-toxins cloned into microbe hosts.  
 XX Disclosure; : 42pp; English.  
 XX DNA encoding the insecticidal toxin can be cloned into baculo  
 CC viruses and transferred to other host microbes, preferably E. coli  
 CC N4522(PMYC 2316) NRRL B-18785, to control acaride pests, or to  
 CC plants which become resistant to the acaricide pests.  
 CC Specifically, the two-spotted spider mite (Tetranychus urticae) is  
 CC controlled.  
 XX Sequence 1257 AA;  
 SQ

Query Match 4.3%; Score 107; DB 14; Length 1257;  
 Best Local Similarity 19.3%; Pred. No. 1.4;  
 Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LITGLGLASLG-LASPVTSYTSVREAPFGYKPGKSEIENLKDKNIVWLLEN--- 61  
 DB 74 vlsvltlpevtgtsaastivsfwfkfgdkpnaknifeelkpglealigqditnyqd 133  
 QY 62 ---RSFNLGGVRRGGLDNPNGPCNFKNASDPSSGKCYCTQAKDYVSFND--- 112  
 DB 134 alnqkffdsigktlnlytv--aldndyvtak-----tqlenlnsiltssdisif 180  
 QY 113 -PDHVSVTGNLEFGYTPNNGAIASGVV-ADQSGFLNAQLNDYPKLAPEEATROVMGY 170  
 DB 181 ipegyetyg-gipyamvanahillirdaivnaeklgfskdevthkky-----ikntlm 234  
 QY 171 YTEESVPTLVLDVDETFP--NSWESCVPGPVTPNRLCALAGTAGHGKNDDFLNYGIS 228  
 DB 235 hteavikafingldkfksldvnsy-----nknany--- 264  
 QY 229 SKSIFEANEKGVSWNLNGTNGEPEPDSLFTTYVN-----QTSRSNVVVENF 277  
 DB 265 ikgmtcmvldivalwptfdpdyhkykeveieftrtlssplygvpkmmqntsssvpsdlf 324  
 QY 278 FQDAYLGLVLPK--FSS-----YINPSCGT-----NNSXHPGTGN 309  
 DB 325 h--yqgdlvklefstrtdnglakiftgtrntfkykspnthethydvfsyntqs---sgn 378  
 QY 310 VSYG-----EVFKQIYD-----AIRGQPQWDKTLF 336  
 , , , ,

DB 379 isrgssnipidlnnpilistcirnsfykalagssvlnfkdggtggyafaqap----- 430  
 QY 337 ITYDETGFGYDH--VPPPLAVRPDNLTYTETAKNG-----QKYTLHFDRLGGRMPTWVI 388  
 DB 431 -----tggawdhshfiesdgapeghklnylytspgdtlrdrfinvytliisptinelsteki 485  
 QY 389 SPY-SKKGYIEQYGTDPVTGKP 409  
 DB 486 kgfpaekgyiknqgimkyygkp 507

RESULT 12  
 WI3886  
 ID WI3886 standard; protein; 1257 AA.  
 XX AC WI3886;  
 XX DT 14-MAY-1997 (first entry)  
 XX DE 33F2 toxin.  
 XX KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;  
 KW biological control; Monomorium pharaonis; delta-endotoxin; lepidoptera;  
 KW insect.  
 XX OS Bacillus thuringiensis isolate PS33F2.  
 XX PN US5596071-A.  
 XX PD 21-JAN-1997.  
 XX PF 22-MAY-1991; 91US-0703977.  
 XX PR 24-NOV-1993; 93US-0158232.  
 PR 22-MAY-1991; 91US-0703977.  
 PR 25-NOV-1991; 91US-0797645.  
 PR 22-MAY-1992; 92US-0887980.  
 XX (MYCO ) MYCOGEN CORP.  
 XX Foncerrada L, Fu J, Kennedy MK, Meier H, Payne JM;  
 PI Randall JB, Schnepf HE, Schwab GE, Uick HJ;  
 XX WPI: 1997-107615/10.  
 DR N-PSDB; T60071.  
 XX Bacillus thuringiensis toxin - active against hymenopteran pests  
 PT Disclosure; Column 57-64; 64pp; English.  
 XX This sequence represents the 33F2 toxin isolated from the Bacillus  
 CC thuringiensis (B.t.) isolate PS33F2. B.t. is a gram-positive, spore  
 CC forming, soil bacterium, characterised by parasporal crystalline protein  
 CC inclusions. These proteins can be highly toxic to pests, and have been  
 CC used to produce insect resistant plants. The previously isolated B.t.  
 CC delta-endotoxins were mainly active against lepidopteran insects, however  
 CC the proteins of the invention are active against hymenopteran insects.  
 CC This sequence is an example of a toxin of the invention, for which the  
 CC sequences shown in WI3886 and WI3871 represent the generic formulae. As  
 CC the toxins of the invention are active against hymenopteran pests, they  
 CC can be used for the biological control of ants, particularly pharaoh ants  
 CC (Monomorium pharaonis).  
 XX Sequence 1257 AA;  
 SQ

Query Match 4.3%; Score 107; DB 18; Length 1257;  
 Best Local Similarity 19.3%; Pred. No. 1.4;  
 Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LITGLGLASLG-LASPVTSYTSVREAPFGYKPGKSEIENLKDKNIVWLLEN--- 61  
 DB 74 vlsvltlpevtgtsaastivsfwfkfgdkpnaknifeelkpglealigqditnyqd 133

QY 62 ----RSPDNLILGVRROGLDNPINNGPCFNKYNKASDPSSGKYCTQAKDYDSVFND----- 112  
 Db 134 ainqkksdlsqtkinlytv--aidnddyvtak-----tqlenlnsiltssdisif 180  
 QY 113 -PDHVSVTGNLLEFGYTYTPNNGATASGVV-ADQSGFLNAQLNDYPKLAPEEATROVMGY 170  
 Db 181 lpegdyetg-glpvyamvanahillirdaivnaeklgfsdkevdkhky-----lkmthm 234  
 QY 171 YTEEEVPTLVDLDEFTTF--NSWFSCVPGPTNPRLCALAGTAAGHGKNDDELNYGIS 228  
 Db 235 hteavikafingldkfksldvnsy-----nkany--- 264  
 QY 229 SKSIFEAAKEGVSWLNVDCTNGEPEPDSLFFTYVN-----QTSRSNVVPVENF 277  
 Db 265 lkgmtewldlvalwptfdpdykveleftrtissplyqvpkmmqntsssiypsdlf 324  
 QY 278 FODAYLGLVLPK--FS-----YINPSCCGT-----NTNSMHPTGN 309  
 Db 325 h---yggdlvklefstrtdndglakiftgirtfkykspnthetvhdvfyntqs---sgn 378  
 QY 310 VSVG-----EVEVKQIYD-----AIROGPQWDKTLIF 336  
 Db 379 isrgssnpiidlnpiistcirsfykaagssvlvnfkdgtqgyafaqap----- 430  
 QY 337 ITYDETGGFYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLHFDRLGGRMPTWVI 388  
 Db 431 ----tggawdhstfiesdgapeghklnlytspgdtlrdfinvytlstptinelstekl 485  
 QY 389 SPY-SKKGYIEQYGTDPVTKGP 409  
 Db 486 kgfpaekgyikngimkyygkp 507  
 RESULT 13  
 R20068  
 ID R20068 standard; Protein; 1257 AA.  
 AC R20068;  
 XX 26-MAR-1992 (first entry)  
 DT B.thuringiensis toxin PS33F2.  
 DE worm; nematocide; flukicide; anthelmintic; parasite; ss.  
 KW Bacillus thuringiensis.  
 XX EP462721-A.  
 XX 27-DEC-1991.  
 XX 04-JUN-1991; 91EP-0305047.  
 XX 03-MAY-1991; 91US-0693018.  
 PR 11-JUN-1990; 90US-0535810.  
 PR 24-JUL-1990; 90US-0557246.  
 PR 27-JUL-1990; 90US-0558738.  
 PR 10-AUG-1990; 90US-0565544.  
 PR 14-MAR-1991; 91US-0669126.  
 PR 27-MAR-1991; 91US-0675772.  
 XX (MYCO-) MYCOGEN CORP.  
 XX Narva KE, Payne JM, Schwab GE, Hickie LA, Galasan T;  
 PI Sick AJ;  
 XX WPI; 1992-001086/01.  
 DR N-PSDB; Q20336.  
 XX New Bacillus thuringiensis strains expressing toxins - have  
 PT nematocidal activity, to control nematodes, helminths and flukes  
 PT e.g. liver fluke Fasciola hepatica

Claim 6; Page 37; 47pp; English.  
 The PS33F2 toxin gene was isolated from a B.t. strain PS33F2 gene library after hybridisation to a 1.8kb DNA fragment. (The probe was obtained by PCR amplification of a 5.85kb fragment which hybridised to probes based on the N-terminal sequence of PS33F2 toxin). The gene was subcloned in a vector to give plasmid pMVC2316 which was transformed into the acrySTALLiferous B.t. strain, HD-1 cryB. The recombinant B.t. strain was grown to sporulation and the toxin protein purified by NaBr gradient centrifugation.  
 See Q20334-Q20343.  
 Sequence 1257 AA:  
 Query Match 4.2%; Score 106; DB 13; Length 1257;  
 Best Local Similarity 19.3%; Pred. No. 1.7;  
 Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;  
 QY 6 LUTGLGLLASLG-LASPVTSYTSVREAPFGYKPGSKESIENLKDKNVENIWLLEN--- 61  
 Db 74 vlsvltlpevgtvsaasativsfwlpkifgdkpnaknifeelkpqlealigqditnysd 133  
 QY 62 ----RSPDNLILGVRROGLDNPINNGPCFNKYNKASDPSSGKYCTQAKDYDSVFND----- 112  
 Db 134 ainqkksdlsqtkinlytv--aidnddyvtak-----tqlenlnsiltssdisif 180  
 QY 113 -PDHVSVTGNLLEFGYTYTPNNGATASGVV-ADQSGFLNAQLNDYPKLAPEEATROVMGY 170  
 Db 181 lpegdyetg-glpvyamvanahillirdaivnaeklgfsdkevdkhky-----lkmthm 234  
 QY 171 YTEEEVPTLVDLDEFTTF--NSWFSCVPGPTNPRLCALAGTAAGHGKNDDELNYGIS 228  
 Db 235 hteavikafingldkfksldvnsy-----nkany--- 264  
 QY 229 SKSIFEAAKEGVSWLNVDCTNGEPEPDSLFFTYVN-----QTSRSNVVPVENF 277  
 Db 265 lkgmtewldlvalwptfdpdykveleftrtissplyqvpkmmqntsssiypsdlf 324  
 QY 278 FODAYLGLVLPK--FS-----YINPSCCGT-----NTNSMHPTGN 309  
 Db 325 h---yggdlvklefstrtdndglakiftgirtfkykspnthetvhdvfyntqs---sgn 378  
 QY 310 VSVG-----EVEVKQIYD-----AIROGPQWDKTLIF 336  
 Db 379 isrgssnpiidlnpiistcirsfykaagssvlvnfkdgtqgyafaqap----- 430  
 QY 337 ITYDETGGFYDH--VPPPLAVRPDNLTYTETAKNG-----QKYLHFDRLGGRMPTWVI 388  
 Db 431 ----tggawdhstfiesdgapeghklnlytspgdtlrdfinvytlstptinelstekl 485  
 QY 389 SPY-SKKGYIEQYGTDPVTKGP 409  
 Db 486 kgfpaekgyikngimkyygkp 507  
 RESULT 14  
 Y29830  
 ID Y29830 standard; Protein; 1041 AA.  
 XX Y29830;  
 AC Y29830;  
 XX 15-NOV-1999 (first entry)  
 DT Saccharomyces cerevisiae ATH1 protein.  
 DE Saccharomyces cerevisiae; baker's yeast; neutral trehalase; NTH1;  
 KW drying-resistant; ATH1; acidic trehalase; desiccation; bread.  
 XX Saccharomyces cerevisiae.  
 OS EP945511-A2.  
 PN

XX 29-SEP-1999.  
 XX PD  
 XX PF 24-FEB-1999; 99EP-0103560.  
 XX PR  
 XX PR 04-MAR-1998; 98JP-0067672.  
 XX (NORQ ) NAT FOOD RES INST MIN AGRIC.  
 XX (ORIY ) ORIENTAL YEAST CO LTD.  
 XX  
 XX Mori K, Nakajima R, Shima J, Suzuki Y, Takano H;  
 XX Watanabe H;  
 XX WPI; 1999-530049/45.  
 XX DR N-PSDB; 207051.  
 XX  
 XX New dried baker's yeast, with greater resistance to desiccation  
 XX  
 XX Disclosure; Page 34-38; 56pp; English.  
 XX  
 CC The present invention describes a diploid dry baker's yeast (I) which  
 CC is obtained by mating an a-type haploid yeast whose ATH1 (acidic  
 CC trehalase gene) or NTH1 (neutral trehalase gene) gene is disrupted with  
 CC an alpha-type haploid yeast disrupted in the same gene. Both haploid  
 CC strains are derived from practical baker's yeast. The present sequence  
 CC represents ATH1 from baker's yeast (Saccharomyces cerevisiae). Yeast  
 CC requires trehalose during rehydration after a period of desiccation.  
 CC (I) is mutated in either the ATH1 or NTH1 gene which increases  
 CC intracellular trehalose and is derived from parent baker's yeast  
 CC strains which already have resistance to drying. Therefore (I) has  
 CC greater resistance to desiccation.  
 XX  
 XX Sequence 1041 AA;

Query Match 4.2%; Score 105.5; DB 20; Length 1041;  
 Best Local Similarity 19.7%; Pred. No. 1.4;  
 Matches 89; Conservative 64; Mismatches 140; Indels 159; Gaps 25;

QY 22 VTSEY---TSVREAPFGYKPGSKESIENLKQVENIVMLILENRS-----PDNLTGG 70  
 DB 446 vdyehlnvdvamasfsiylnghgidd--eylyttwpliknaaqftayvknyslgl 503  
 QY 71 VRQGLDNP-----INNGPCNYKNASDPSSGKYCTQAKDYDSVFNDPDHSVTGNLLEF 124  
 DB 504 yetynltddpdefanhmgaf---tnagiktlkwatdgnhlgevdpkwselskdi-- 558  
 QY 125 YGYTPNNGAIGASKVVDQSGFLNAQLNDYPKLAPEATROV--MGYTTREEVPTLVLDL 182  
 DB 559 ---yipr-----ssnitleysg-----mnssvelkqadvlmyvplgyindesi--lnna 604  
 QY 183 VDEFTTFNSWFSCV-PGPTNPRLCALAGTAAGHG-----K 217  
 DB 605 ikdlyyyeersgaspamtypvfvaaag-llnhgssqsylyksvlpylrapafsqe 663  
 QY 218 NDDDFLNYGISKSI-FEANEK-----GVSWLNYDGTNGEFPDPSLFTVYNOTSS 269  
 DB 664 sddnflngltqapflitanggfllqilgtlgyr---syevdpdt---kkinrlrf 717  
 QY 270 NVVPVENFQDAYLGVLP-----KFSVINPSCCGTNTSMHPTGNVSYGEVVKQIYA 323  
 DB 718 n--pie-----lplpggiainrfkymn-----qvid- 742  
 QY 324 IRQGFQWDKTLFTYVDETFGGFYDH-----VPPPLAVRPDNLTYTETAKNGOKYT 373  
 DB 743 -----fiidhngitvkhsgdvpilkiprslilhdgdnfngyngenerkpn 789  
 QY 374 LH---FDRLGGRMPTWISPYSKGYTEOYGT 402  
 DB 790 leardvdvrgdpm-----rmdrygt 809

RESULT 15

R64927  
 ID R64927 standard; Protein; 1122 AA.  
 XX AC R64927;  
 XX DT 10-AUG-1995 (first entry)  
 XX DE Cytadhesin protein.  
 XX KW Cytadhesin protein; immunoassay; antigen.  
 XX OS Mycoplasma gallisepticum.  
 XX PN US5378820-A.  
 XX PD 03-JAN-1995.  
 XX PF 09-NOV-1992; 92US-0973257.  
 XX PR 09-NOV-1992; 92US-0973257.  
 XX PA (DOHM/) DOHMS J E.  
 XX PA (KEEL/) KEELER C L.  
 XX PI Dohms JE, Keeler CL;  
 XX DR WPI: 1995-051314/07.  
 XX DR N-PSDB; Q81778.  
 XX PT Nucleic acid encoding cytoadhesin protein - used as a probe to  
 XX PT diagnose Mycoplasma gallisepticum infection in poultry  
 XX PS Claim 1; Column 7-16; 13pp; English.  
 XX  
 CC The protein is a cytoadhesin protein from Mycoplasma gallisepticum,  
 CC and may be used as a diagnostic antigen (optionally in recombinant  
 CC form) in immunoassay formats for diagnosis of M. gallisepticum  
 CC infection in poultry, e.g. turkey or fowl.  
 XX  
 XX Sequence 1122 AA;

Query Match 4.2%; Score 105.5; DB 16; Length 1122;  
 Best Local Similarity 20.4%; Pred. No. 1.6;  
 Matches 134; Conservative 63; Mismatches 209; Indels 251; Gaps 33;

QY 3 STALLTGLGLLASIGLASPVTS-----EYTS-VREAPFGYKPGSKESIENLKQVENIV 55  
 DB 216 starliveegnsagmgtmtptastsetvidynsdnkip---kpktkdsessesing-- 270  
 QY 56 WLILENRFDMILGVRROGL-----DNP--INNGPCNYKNASD--PS----- 95  
 DB 271 -----grtyanintqknkgvkvnenkfnsenpfavenmafikpkmvmdnypstwtgg 325  
 QY 96 --SGKYCT-----QAKDY-----DSVFNDPDHSVTGNL 122  
 DB 326 sangkmtvkvfykhdnpnavnrfakyyprkrltqttdlidsfsfyehepweygn 385  
 QY 123 EFY-----GTYTPNNGAIGASG-----K 139  
 DB 386 qfvmppmqiytknglyakdgmvyfngngtwvnnesalsigvfrtkferntaeapgnk 445  
 QY 140 VVADQSGLF-----NAQLNDYPKLAPEATROVMGYTTEEVPTLVLDVDEFTTFNSWF 193  
 DB 446 tvgyypgilllsaisdatnglalaap--akgqdvgyhf--vprl----- 486  
 QY 194 SCVPGPTNPRLCA--LAGTAAGHGKNDDDFLNYGISKSIFFEA-- 236  
 DB 487 -avgvssprgngniflgsaltwtngngnfltkwhspaviedaptftitvnssgvln 545  
 QY 237 -----NEKGVSWLNYDGTNGEFPDPSLFTVYNOTSSRVVPEVFODAY-- 282  
 DB 546 sgsqststpmppnsngnesipyrtwnsydynsvrfaaliskpagntkqveskftail 605







120 snAsnLeuGluPheTyrglyThrThrProAsnAsnGlyAlaIleAla 136  
137 SerGlyLysValAlaIleAspGlnSerGlyPheLeuAsnAlaGlnLeuAs 153  
1011 AGTGGCAGGTCGTGGCGACCGTCTGCTTCTCCTCAACGCACAGCTTAA 1060  
153 nAspTyProLysLeuAlaProGluAlaThrArgGlnValMetGlyT 170  
1061 CGACTACCCCAAACTGGCCCGACAGAGCGGACAGGCAAGTGTGGAT 1110  
170 yTrThrGluGluValProThrLeuValAspLeuValAspGluPhe 186  
1111 ACTATACCGAGGAGGAGTCTACGCTCGTGACCTTCTGGATGAGTTC 1160  
187 ThrThrPheAsnSerTrpPheSerCysValProGly..... 198  
1161 ACTACTTCAACAGCTGTCTCTGCTGTCTTCTGGCGTAAGTGGGATAA 1210  
199 .....ProThrAsnProA 203  
1211 TCTGTATATTTTAAATTTTGACTAACGCGTTCAAAAGGCTTACCAACCCCA 1260  
203 snArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsnAsp 219  
1261 ACCGCTGTGGCTCTGGCAGAACCGCTGCTGGCGATGGCAAGATGAC 1310  
220 AspAspPheLeuAsnTyrglyIleSerSerLysSerIlePheGluAlaAl 236  
1311 GATGACTTCTGAATATGATATCTAGCAAGTCCATCTCGAGGCGC 1360  
236 aAsnGluLysGlyValSerTrpLeuAsnTyrglyThrAsnGlyLup 253  
1361 CAACGAGAGGGGCTGCTGCTGCTCAACTACGATGGCACCACGAGAT 1410  
253 heGluProAspSerLeuPhePheThrTrpValAsnGlnThrSerArgSer 269  
1411 TCGAACCGGATCTCTCTCTCTACCTACCTCAACGACAGCTCCGCGTCC 1460  
270 AsnValValProValGluAsnPheGlnAspAlaTyrluGlyValle 286  
1461 AACGTGTCGCGCTTGAAGAACTTCTTCCAAAGCGCTACCTCGGTGCT 1510  
286 upToLysPheSerTyrlleAsnProSerCysGlyThrAsnThrAsn 303  
1511 CCTAAATTTCTTACATTAACCCCTCTGCTGGCGCACCACCAACT 1560  
303 erMetHisProThrGlyAsnValSerTyrglyGluValPheValLysGln 319  
1561 CCATGACCCCGACCGGTACGCTCTCTACGCTGAGGTCTTCTCAACGAG 1610  
320 IleTyAspAlaIleArgGlnGlyProGlnTrpAspLysThrLeuLeuph 336  
1611 ATCTATGATGCCATTGCGCAGGCGCTCAGTGGACAGACCTCTGCTTT 1660  
336 eIleThrTyAspGluThrGlyGlyPheTyrglyAspHisValProProL 353  
1661 CATTAACCTACGAGAGCGGTGCTTCTACGACCATGCTCCCTCCCGCTC 1710  
353 euAlaValArgProAspAsnLeuThrTyrlleGluGlnTyrglyThrA 369  
1711 TCGCGCTCGCGCGGACCAACTGACCTACACTGAGATGCGAAGAACGCT 1760  
370 GlnLysTyThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 386  
1761 CAGAAATACCTTCTACTTCGACCGTCTGGGTGGCGCATCGCGACCTG 1810  
386 pValIleSerProTyrlleSerLysGlyTyrlleGluGlnTyrglyThrA 403  
1811 GGTATCTCCCTTACAGTAAGAGGATACATCGACAGTACGGAACGG 1860  
403 spProValThrGlyLysProAlaProTyrlleSerAlaThrSerValLeuLys 419

1861 ATCCCGTCACGGGCAAGCCGCTCCCTACAGTGTACTCCGCTCTCAAG 1910  
420 ThrLeuGlyTyrlleThrAspIleGluAspPheThrProArgValAlaAla 436  
1911 ACTCTCGGATATCTCTGGACATCGAGACTTCACCCCTCGTGTGCCCA 1960  
436 sSerProSerPheAspHisIleGlyThrLeuArgGluAspAla 453  
1961 CTCTCATCTTTCGATCACTGATCGGACGACTTTCGCTGAGGATGCTC 2010  
453 rolleAlaLeuLysThrProHisThrPheSerVal 464  
2011 CTATGCTCAAGACTCCCATACCTTTTCGTA 2045  
seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-470-350B-1

seq\_documentation\_block:  
; Sequence 1, Application US/08470350B  
; Patent No. 5684126  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xiao  
; APPLICANT: Snyder, Solomon H  
; TITLE OF INVENTION: Ebnerlin: A Secreted von Ebner's Gland  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08470,350B  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wolfe, Susan A  
; REGISTRATION NUMBER: 33,568  
; REFERENCE/DOCKET NUMBER: 01107.48790  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4360 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Rattus rattus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 94..3963  
; US-08-470-350B-1

alignment\_scores:  
Quality: 122.00 Length: 303  
Ratio: 0.878 Caps: 16  
Percent Similarity: 45.875 Percent Identity: 23.762  
alignment\_block:  
US-09-426-072-2 x US-08-470-350B-1



us-09-426-072-2.rni

Tue Feb 27, 08:39:04 2001

3493 ..TTCTGCTCTTTGACATATGATCTCTCATCAGACTGGATGCATACAGAT 3540

Align seg 1/1 to: US-08-470-350B-1 from: 1 to: 4360

```

86 CysAsnTyr.....LysAsnAlaSerAspPr 94
|||||
2758 TGCACATATGACTATATAGATTTTGTATGGCCCCCACCACAGTTTACC 2807
94 oSerSerGlyLysTyrCys.....ThrGlnAlaL 104
|||||
2808 TCTCATTTCCCGGGTTGTGTATGGGCCATGGCTCTTTCACCTTCAACAT 2857
104 ysAspTyrAspSerValPheAsnAspProAspHisSerValThr..... 118
|||||
2858 CCAACTTCATGTCAGTTCGGTTCACCACTGATCAGTGTACTTCAAGA 2907
119 GlyAsnAsnLeuGluPheTyrGlyThrProAsnAsnGlyAlaI 135
|||||
2908 GGGTTCGGGCTGACTACTACTGACGTTTGACATAATATACCAATCT 2957
135 eAlaSerGlyLysValAlaAlaAspGlnSerGlyPheLeuAsnAlaGlnL 152
|||||
2958 CCTTGT.....C 2965
152 euAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnValMet 168
|||||
2966 TGTCAATACATGAGAGCCAGTGTGAGCAGGAGTACCTTTCAGTCCATG 3015
169 GlyTyrThrGluGluValProThrLeuValAspGluVal 185
|||||
3016 GCCTACTCTCCAGG.....GATCTTGT 3039
185 upHeThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn 202
|||||
3040 ..ATCTCTGGTGGACGTCAGTTCACAGTGTGAGGCTCAGATAACAA 3088
202 toAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsn 218
|||||
3089 GGGAGTTCATATTCACATTCCTTACACAGGTCGGTACTACCAACAG 3138
219 ...AspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePheG 234
|||||
3139 GGTGACACAGCAGCAGTAC.....TCCAACTTCTCTCAA 3176
234 uAlaAlaAsnGluLysGlyVal..... 241
3177 ACGGCTGTTTCAATGGCATCATCAAAAGGAGAAAGATCTCCACATCC 3226
242 .....SerTrpLeuAsn.....TyrAsp 247
3227 ATGTGAGTCAAGATGCTTTCAGAACACCTGGGTCAACACCATGATCATC 3276
248 GlyThrAsn.....GlyGluPheGluPr 255
|||||
3277 ACCAACACACAGTCGAGATCCAGAGTCCAGTATGCAATTTTGAGT 3326
255 oAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSerAsnVal 272
|||||
3327 GAATATTTCTTTTAT.....ACATCTCTCTCTCTTGT 3361
272 alProValGlu.....AsnPheGlnAspAlaTyr 282
|||||
3362 ATTCAGTGCACGAGCCCATATTTATGTTGGATCTGGACCAATTTGTAC 3411
283 Leu..GlyValLeuProLysPheSerTyrIleAsnProSerCysGly 298
|||||
3412 CTTTACGGCGAGTCTTCCATTCGATACCTCTTTGGCTCTGTTGTGGA 3461
299 ..ThrAsnThrAsnSerMetHisProThrGlyAsnValSerTyrGlyGlu 315
|||||
3462 CACCTGTGTGGCTTCGCCACATCCCAATGAC..... 3492
315 alPheValLysGlnIleTyrAspAlaIleArgGlnGlyProGlnTrpAsp 331
|||||

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-494-714-1

```

seq_documentation_block:
; Sequence 1, Application US/08494714
; Patent No. 5587290
; GENERAL INFORMATION:
; APPLICANT: Klionsky, Daniel
; APPLICANT: Holzer, Helmut
; APPLICANT: Destrielle, Monica
; TITLE OF INVENTION: STRESS TOLERANT YEAST MUTANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494.714
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard Aron
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-61036/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 210 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3876 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 579..3701
; US-08-494-714-1

```

```

alignment_scores:
Quality: 111.50      Length: 446
Ratio: 0.555        Gaps: 25
Percent Similarity: 45.067      Percent Identity: 19.955

alignment_block:
US-09-426-072-2 x US-08-494-714-1 ..

Align seg 1/1 to: US-08-494-714-1 from: 1 to: 3876

22 ValThrSerGluTyr.....ThrSerValArgGluAlaProPheG 35
|||||
1914 GTGATTACGAATACCATATTAACGTTGATGCTGCTATGCGCTCTCTTTC 1963

35 yTyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysVal 52
|||||
1964 CATATCTGTGATGGACACGAGGATTCATGAC.....GAGTATCTGA 2007

52 luAsnIleValTrpLeuIleLeuGluAsnArgSer..... 63

```

```

2803 ATATAATAATGACGACCACCAATGGTGACGATTGTTCTATAAATCAGGAGAT 2852
|||||
310 ValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgGI 326
|||||
2853 GTT .....CCTATTCMTAATAG ..... 2870
326 nGlyProGlnTTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrG 343
2870 ..... 2870
343 LyGlyPheTyrAspHisValProProLeuAlaValArgProAspAsn 359
|||||
2871 .....ATACCAAAACAGATCTCTAATACATCACCAGGAT 2903
360 LeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHis.. 375
|||||
2904 ATCAACTCTATATGTTCCGAAACGAAGAAGAACCAATCTAGAGCG 2955
376 .....PheAspArgLeuGlyGlyArgMetProThrTrpValIleSerP 390
2954 TAGAGAGCTGCACCGTGTTGGTGATCCAAATG ..... 2984
390 roTyrSerLysGlyTyrIleGluGlnTyrGlyThr 402
2985 .....AGGATGGATAGGTATGGTACC 3005
seq_name : /cgn2_6/prodata/2/Ina/PCTus_COMB.seq;PCT-US96-10782-1
seq_documentation_block :
; Sequence 1, Application PC/TUS9610782
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University
; APPLICANT: of California
; TITLE OF INVENTION: STRESS TOLERANT YEAST MUTANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBBINS, BERLINER & CARSON
; STREET: 201 N. Figueroa Street, 5th Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90012-2628
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10782
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Berliner, Robert
; REGISTRATION NUMBER: 20,121
; REFERENCE/DOCKET NUMBER: 5555-400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 977-1001
; TELEFAX: (213) 977-1003
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3876 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 579..3701
; ORIGIN: 1
; TT-US96-10782-1

```

alignment\_scores:

Quality: 111.50 Length: 446  
Ratio: 0.555 Gaps: 25  
Percent Similarity: 45.067 Percent Identity: 19.955

## alignment\_block:

US-09-426-072-2 x PCT-US96-10782-1 ..

Align seg 1/1 to: PCT-US96-10782-1 from: 1 to: 3876

```

22 ValThrSerGluTyr.....ThrSerValArgGluAlaProPheG1 35
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1914 GTCGATTACGAATACCATATTAAAGTTGATGCTGCTATGCGCTCTCTTTC 1963
35 yTyrLysProGlySerLysGluSerLleGluAsnLeuLysAspLysValG 52
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
1964 CATATACVTGAATGACACGACGAAGGATTGATGAC.....GAGTATCTGA 2007
52 LuAsnIleValTrpLeuIleLeuGluAsnArgSer..... 63
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2008 GATATACTACATGCCAATTATCAAAAACGACGCCCAATTTTACTGCT 2057
64 .....PheAspAsnIleLeuGlyClyValArgGlnGlyLeuAs 77
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2058 TATGTAAGTACATATCTCCCTAGGATTGTATGAACACATATAATTTCAC 2107
77 pAsnPro.....IleAsnAsnGlyProPheCysAsn 88
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2108 AGATCCCGCAGTTGCTAAATCACATCAATAACGGGGCTTTC..... 2150
88 yLysAsnAlaSerAspProSerSerGlyLysTyrCysThrGlnAlaLys 104
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2151 ..ACGAATGCTGGCATTAACACACTCTTAAGTGGGACACACATTCGGC 2198
105 AspTyrAspSerValPheAsnAspProAspHisSerValThrGlyAsnAs 121
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2199 AATCATCTCGCGAGGTCGTGTGACCCCAATGGAGTGAATTTCCAAAGA 2248
121 nLeuGluPheTyrGlyThrTyrProAsnAsnGlyAlaIleAlaSerG 138
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2249 TATT.....TATATCCCTAGA.....TCCTCAT 2271
138 yLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAsp 154
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2272 CTAAATCATCTTTGGAAATATCTGCT.....ATGAATAGC 2306
155 TyrProLysLeuAlaProGluGluAlaThrArgGlnVal.....MetG1 169
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2307 TCAGTCGAGATTAAACAGCGGATGTGACTTTAATGTTTACCCACTTGG 2356
169 yTyrTyrThrGluGluGluValProThrLeuValAspLeuValAspGluP 186
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2357 ATATATCAATGATGAATCCATA.....TTGAACAATGCAATTAAGATC 2400
186 heThrThrPheAsnSerTrpPheSerCysVal...ProGlyProThrAsn 201
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2401 TTTATATTATTTCAGAAAGACAGCTGCTGCTGGCCTGCAATGACATAT 2450
202 ProAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGly..... 216
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2451 CCGGTTTTTGTGGCGCAGCTGTGGT....CTGCTGAATCAGCGCTCTTC 2497
216 ..... 216
2498 TTCTCAAAGTTACTTATATAATCGGTTCTTCCATACCTACGGGCTCCTT 2547
217 .....LysAsnAspAspPheLeuAsnTyrGlyIle 227
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2548 TCGTCAATTTAGTGACGCAATCAGACGCACTTTTACACACGAGATTA 2597
228 SerSerLysSerIle...PheGluAlaAlaAsnGluLys..... 239
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2598 ACCCAGCCAGCATTCCTCCCTTTTAAACAGCTAACGGTGGGATTTCTACAGAG 2647

```

```

240 .....GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluP 253
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2648 CATTCCTGTTGGGTTAAACAGGAATCCGATAT.....TCATTATGAGG 2688
253 heGluProAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSer 269
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2689 TTGATCCAGATCT.....AAAAAATTAACCGTTTGTGTAAGGTTTC 2729
270 AsnValValProValGluAsnPhePheGlnAspAlaTyrLeuGlyValLe 286
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2730 AAT.....CAAATAGAA.....CTACCGTTGCT 2752
286 uPro.....LysPheSerTyrIleAsnPro..... 294
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2753 CCCTCGTGGTATCGCTATTAGAAACTTCAAAATATATGAACCCAGCTTTAG 2802
295 .....SerCysCysGlyThrAsnThrAsnSerMetHisProThrGlyAsn 309
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2803 ATATATTAATTTGACGACCAATGTCAGGATTGTTCAATAATCAGAGAT 2852
310 ValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgG1 326
   ||| : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2853 GTT.....CCTATTCTATATAAG..... 2870
326 nGlyProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrG 343
2870 ..... 2870
343 lyGlyPheTyrAspHisValProProLeuAlaValArgProAspAsn 359
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2871 .....ATACCAACACAGATCTCTAATACATGACCAAGAT 2903
360 LeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHis... 375
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2904 ATCAACTCTATATAATGTTCCGAAACGAAAGAAACCAAACTAGAGCG 2953
376 .....PheAspArgLeuGlyClyArgMetProThrTrpValIleSerP 390
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2954 TAGAGACGTCGACCGCTGTTGGTGATCCAATG..... 2984
390 roTyrSerLysLysGlyTyrIleGluGlnTyrGlyThr 402
   : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
2985 .....AGGATGGATAGGTATGTTGATGCTAC 3005
seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5223424-3
seq_documentation_block:
; Patent No. 5223424
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,
; RICHARD D.
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
; HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
; ACID SEQUENCE
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/225,032
; FILING DATE: 27-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 78,519
; FILING DATE: 27-JUL-1987
; APPLICATION NUMBER: 933,107
; FILING DATE: 20-NOV-1986
; APPLICATION NUMBER: 902,887
; FILING DATE: 02-SEP-1986
; APPLICATION NUMBER: 887,140
; FILING DATE: 17-JUL-1986
; APPLICATION NUMBER: 823,102
; FILING DATE: 27-JAN-1986
; APPLICATION NUMBER: 773,430
; FILING DATE: 06-SEP-1985
; SEQ ID NO: 3:
; LENGTH: 2062
; 5223424-3

```

```

alignment_scores:
  Quality: 109.50
  Ratio: 0.622
  Percent Similarity: 43.781
  Percent Identity: 21.393
alignment_block:
  US-09-426-072-2 x 5223424-3
  Align seg 1/1 to: 5223424-3 from: 1 to: 2062
  110 PheAsnAspProAspHisSerValThrGlyAsnAsnLeu..... 122
  627 TTCAATCCAGCGGACTGGCAGTAAATATCCCAACAACATGCAGAAATAAA 676
  123 .....GlupHeTyr..... 125
  677 CTTAGTAGTTTTCACAGAAATATTCAATGCTAGTACTTAACACATTA 726
  126 .....GlyThyrThrPro.....AsnAsnGlyAlaIle 135
  727 CAGAAATCAGCAACCTCCACCACCTCCAAATATATATATGATCTTAAC 776
  136 AlaSerGlyLysValAlaAspGlnSerGlyPheLeuAsnAlaGlnLe 152
  777 GCAAGCTTAATGCTCGCACTAGACCAATACACACTT..... 815
  152 uAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnValMetG 169
  816 .....CNAACACACCGACGACCTAGAAGTGAACA.....CTTG 852
  169 lYtYrThrGluGluGluValProThrLeuValAspLeuValAspGlu 185
  853 GTTTTAT.....CCATGCTTACTACA.....AAA 878
  186 PheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsnPr 202
  879 CCAACTCAATACAGATATTACCTATCATGCTCAGAAACCTTAATCCACC 928
  202 oAsnArgLeuCysAlaLeuAlaGlyThrAlaGlyHisClyLysAsnA 219
  929 A.....ACATACACTGGCAATCATCACACAAA 954
  219 sPAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePheGluAla 235
  955 TAACAGACTCAATACAAACAGGACTACAGATGACATATGTTCTACACA 1004
  236 AlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGly.....Th 249
  1005 ATAGAAATGCAGTACCAATTCATCTTCTTGAAGACTGGAGATGAATCTC 1054
  249 rAsnGly.....GlupHeGluProAspSerLeuPhePheThrTyrValA 264
  1055 CAGAGAAATATATCACTTTGCACAAAACCAATTAATTAATCACTCAT 1104
  264 snGlnThrSerArgSerAsnValValProValGluAsnPhePheGln... 279
  1105 GGCAAAACAACAGATCTCTAGAGCTGGCTCCAAAAGTACTAACTGAAC 1154
  280 .....AspAlaTyrLeuGlyValLeuProLysPheSerTyrIl 292
  1155 ACCACAGAACGACACCAACCCAGGAACACTACCA..... 1190
  292 eAsnProSerCysCysGlyThrAsnThr...AsnSerMetHisProThrG 308
  1191 .....GGAGCTACACACAGAAAGGTTATCAACCAACA 1224
  308 lYAsnValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIle 324
  1225 TTAATAATAGCTACACAGAACG.....ACAGCACT 1256
  325 ArgGlnClyPro...GlnTrpAspLysThrLeuLeuPheIleThrTyrAs 340

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-876-280-5

```

seq_documentation_block:
/ Sequence 5, Application US/07876280
/ Patent No. 5262158
/ GENERAL INFORMATION:
/ APPLICANT: Payne, Jewel M.
/ APPLICANT: Cannon, Raymond J.C.
/ APPLICANT: Bagley, Angela L.
/ TITLE OF INVENTION: No. 5262158a1 Bacillus thuringiensis Isolates for
/ NUMBER OF INVENTION: Controlling Acarides
/ NUMBER OF SEQUENCES: 30
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: David R. Saliwanchik
/ STREET: 2421 N.W. 41st Street, Suite A-1
/ CITY: Gainesville
/ STATE: FL
/ COUNTRY: USA
/ ZIP: 32606
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC Compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/876, 280
/ FILING DATE: 19920430
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Saliwanchik, David R.
/ REGISTRATION NUMBER: 31,794
/ REFERENCE/DOCKET NUMBER: M/S 104
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 904-375-8100

```

113 .....ProAspHisSerValThrGlyAsnAsnLeuGluPheT 125  
 531 CTCATATTTATCCAGAGGATATGAACCTGGA...GGTTTACCTTAT 577  
 125 yfGlyThrTyThrProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141  
 578 ATGCTATGTTGCTTAATGCTCATATATTTATTTAAGAGACGCTATAGT 627  
 142 ...AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyProly 157  
 628 AATGCAGAGAAATAGGCTTTAGTGATAGAAGATGACACACATATAAAA 677  
 157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyTyThrGluG 174  
 678 ATAT.....ATCAAAATGACAATACACAATCATATCTAGAG 712  
 174 LuGluValProThrLeuValAspLeuValAspGluPhe..... 186  
 713 CAGTAATAAAGCAATCTTAATGAGCTTGACAAATTTAAGAGCTTTAGAT 762  
 187 ...ThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn... 201  
 763 GTAATAGCTATATAATAAAGCAAAATTTATTAAGGTATGACAGAAAT 812  
 202 .ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217  
 813 GGTCTTGTATCTAGTTGCTCTATGGCCAACTTTTCGATCCAGATCATATC 862  
 217 yAsnAspAspAspPheLeuAsnTyGlyIleSerSerLysSerIlePhe 233  
 863 AAAAGAAGTAGAAATTTGAATTTTACAGAAGCAATTTCTTCCCAATTTAC 912  
 234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyAspGlyThrAs 250  
 913 CAACCTGTACCTAAA.....AACATGCAAAATACCTC 944  
 250 nGlyGluPheGluProAspSerLeuPhe..... 259  
 945 TAGCTCTATTGTACCTAGCTATCTTTTCTACTATCAAGGAGATCTGTAA 994  
 260 .....Phe 260  
 995 AATTAGAAATTTTCTCAAGAAGCGACACAGATGGTCTTGTCAAAATTTT 1044  
 261 ThrTyValAsnGlnThrSerArgSerAsnValProValGluAsnPh 277  
 1045 ACTGGTATTCGAAACACATCTTACAAATCCCTAATACTACTCATGAACATA 1094  
 277 ePheGlnAspAlaTyLeuGlyValLeuProLysPheSerTyIleAsnP 294  
 1095 CCATGTAGAT.....TTTAGTTAT..... 1113  
 294 roSerCysCysGlyThrAsnThrAsnSerMethHisProThrGlyAsnVal 310  
 1114 .....AATACCCAACT.....AGTGGTAATATT 1137  
 311 SerTyGly..... 313  
 1138 TCAAGAGGCTCTCAAAATCCGATTCCAATTGATCTTAATAATPCCATTAT 1187  
 314 .....GluValPheValLysGlnIleTyAsp..... 322  
 1188 TTCAACTTGATAGAAATTCATTTTATAAGCAATAGCGGATCTTCTG 1237  
 323 .....AlaIleArgGlnGly 327  
 1238 TTTTAGTTTAATTTTAAAGATGCGACTCAAGGGTATGCTATTCGCCAAGCA 1287  
 328 ProGlnTrpAspLysThrLeuLeuPheIleThrTyAspGluThrGlyL 344  
 1288 CCA.....ACAGGAGG 1298  
 344 yPheTyAspHis.....ValProProLeuAlaValArgProAspA 359

TELEFAX: 904-372-5800  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3771 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Bacillus thuringiensis  
 INDIVIDUAL ISOLATE: 33f2  
 IMMEDIATE SOURCE:  
 CLONE: E. coli NM522(pMYC 2316) B-18785  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 4..24  
 /function= "oligonucleotide  
 OTHER INFORMATION: hybridization probe"  
 OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
 OTHER INFORMATION: /standard\_name= "probe a"  
 OTHER INFORMATION: /note= "probe A"  
 NAME/KEY: misc\_feature  
 LOCATION: 13..33  
 /function= "oligonucleotide  
 OTHER INFORMATION: hybridization probe"  
 OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
 OTHER INFORMATION: /standard\_name= "Probe B"  
 OTHER INFORMATION: /label= "probe-b"  
 OTHER INFORMATION: /note= "probe b"  
 US-07-876-280-5  
 alignment\_scores:  
 Quality: 107.00 Length: 491  
 Ratio: 0.493 Gaps: 22  
 Percent Similarity: 44.196 Percent Identity: 19.552  
 alignment\_block:  
 US-09-426-072-2 x US-07-876-280-5 ..  
 Align seg 1/1 to: US-07-876-280-5 from: 1 to: 3771  
 6 LeuLeuThrGlyLeuGlyLeuLeuAlaSerLeuGly...LeuAlaSerPr 21  
 220 GTTTATCTCTATTAACTTAATTCCTGAAGTGGTGGTGGCTCTGC 269  
 21 ovalThrSerGluTyThrSerValArgGluAlaProPheGlyTyLysP 38  
 270 AGCAAGTACTATTGTAAAGTTTATTTGGCCCTAAATATTTGGAGATAAC 319  
 38 roGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIle 54  
 320 CAAATGCAAAAAATATATTGAAGAGCTCAAGCTCAAAATTTGAAGCATTA 369  
 55 ValTrpLeuIleLeuGluAsn.....ArgSerPh 64  
 370 ATTCAACAAGATATACAAACTATCAAGATGCAATTAATCAAAAAAAT 419  
 64 eAspAsnIleLeuGlyGlyValArgArgGlnGlyLeuAspAsnProIleA 81  
 420 TGACAGCTCTTCAGAAAAACAATTAATCTATATACAGTA.....GCTATAG 463  
 81 snAsnGlyProPheCysAsnTyTyLysAsnAlaSerAspProSerSerGly 97  
 464 ATAAACAATGATTACGTAAACAGCAAAA..... 489  
 98 LysTyCysThrGlnAlaLysAspTyThrAspSerValPheAsnAsp..... 112  
 490 .....ACGCAACTCGAAAAATCTAAATTTCTATATCTTACCTCAGATAT 530

```

1299 TCGCTGGGACCATCTTTTATTGAATCTGATGGTCCGCCAGAAAGGCATA 1348
      |||
359 snLeuThrTyThrGluThrAlaLysAsnGly..... 369
      |||
1349 AATTAACATATATTACTTCTCCAGGTGATACATTAAGAGATTTCATC 1398
      |||
370 GlnLysTyThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 386
      |||
1399 AATGTATATATACTCTTATAAGTACTCCAACTATAAATGAACATCAACAGA 1448
      |||
386 pValIleSerProTy...SerLysLysGlyTyrlleGluGlnTyGlyT 402
      |||
1449 AAAAATCAAGAGCTTCTCTCGGAAAGGATATATCAAAAATCAAGGGA 1498
      |||
402 hrAspProValThrGlyLysPro 409
      |||
1499 TCATGAATATTACGGTAAACCA 1521

seq_name: /cqn2_6/ptodata/2/ina/5A_COMB.seq:US-08-049-783-1
seq_documentation_block:
; Sequence 1, Application US/08049783
; Patent No. 5439881
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E
; APPLICANT: Schwab, George E
; APPLICANT: Payne, Jewel M
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,783
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1 (PS33P2):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3771 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: 33f2
; IMMEDIATE SOURCE:
; CLONE: 33f2a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4..24
; OTHER INFORMATION: /function= "oligonucleotide
; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"
; OTHER INFORMATION: /standard_name= "probe a"

```

```

; OTHER INFORMATION: /note= "Probe A"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13..33
; OTHER INFORMATION: /function= "oligonucleotide
; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"
; OTHER INFORMATION: /standard_name= "probe B"
; OTHER INFORMATION: /label= probe-b
; OTHER INFORMATION: /note= "probe b"
US-08-049-783-1

alignment_scores:
Quality: 107.00 Length: 491
Ratio: 0.493 Gaps: 22
Percent Similarity: 44.196 Percent Identity: 19.552
alignment_block:
US-09-426-072-2 x US-08-049-783-1
Align seg 1/1 to: US-08-049-783-1 from: 1 to: 3771
6 LeuLeuThrGlyLeuGlyLeuLeuAlaSerLeuGly...LeuAlaSerPr 21
   |||
220 GTTTTATCTGTATTAACCTTTAATTCCTGAAGTTGGTACTGTGCCTCTGC 269
21 oValThrSerGluTyThrSerValArgGluAlaProPheGlyTyrlYsp 38
   |||
270 AGCAGTACTATTGTAAGTTTATTGGCCTAAATATTGGAGATAAAC 319
38 roGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIle 54
   |||
320 CAAATGCAAAAATATATTTCAGAGCTCAAGCCTCAAATTTGAAGCATTA 369
55 ValTrpLeuIleLeuGluAsn.....ArgSerPh 64
   |||
370 ATTCACAAGATATAACAACATCAAGATGCAATTAATCAAAAAAATT 419
64 eAspAsnIleLeuGlyValArgGlnGlnLeuAspAsnProIleA 81
   |||
420 TGACAGCTCTCAGAAACAATTAATCTATATACAGTA.....GCTATAG 463
81 snAsnGlyProPheCysAsnTyrlLysAsnAlaSerAspProSerSergly 97
   |||
464 ATAACAATGATTACGTAAACAGCAAAA..... 489
98 LysTyrcysThrGlnAlaLysAspTyrlAspSerValPheAsnAsp..... 112
   |||
490 .....ACGCAACTCGAAAACTCTAAATCTCTATACCTCAGATAT 530
113 .....ProAspHisSerValThrGlyAsnAsnLeuGluPhe 125
   |||
531 CTCCATATTTATCCAGAGGATATGAAGCTGGA...GGTTTACCTTATT 577
125 yrGlyThrTyThrProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141
   |||
578 ATGCTATGTGCTAATGCTCATATATTATTGTTAAGAGAGCGCTATAGTT 627
142 ...AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrlProly 157
   |||
628 AATGCAGAGAAATTAGCTTTAGTATAAAGAGGTAGACACACATAAAAA 677
157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrlTyThrGluG 174
   |||
678 ATAT.....ATCAAAATGACAATACACAATCATCACTGAAG 712
174 luGluValProThrLeuValAspLeuValAspGluPhe..... 186
   |||
713 CAGTAATAAAGCATCTTAATGAGCTTGACAAATTTAAGAGATTAGAT 762
187 ...ThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn... 201
   |||

```

763 GTAAATAGCTATATAAAAAACCAATTTATTAAGGCTATCAGCAAAAT 812  
202 .ProAsnArgLeuGlyAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217  
813 GGTCTTGATCTAGTGTCTATGCGCACTTTTCGATCCAGATCATTCATC 862  
217 ysAsnAspAspPheLeuAsnTyrGlyIleSerSerIlePhe 233  
863 AAAAAGAGTAGAAATTAATTAACAAGAACTATTCTCTCCCAATTCAC 912  
234 GluAlaAlaAsnGluGlyValSerTrpLeuAsnTyrAspGlyThrAs 250  
913 CAACCTGTACCTAA.....AACATGCAAAATACCTC 944  
250 nGlyGluPheGluProAspSerLeuPhe..... 259  
945 TAGCTCTATTGCTAGCGATCTATTTCACATCAAGGAGATCTTGTA 994  
260 .....Phe 260  
995 AATTAGAAATTTCTACAAGACGACGACGATGCTCTGCAAAATTTT 1044  
261 ThrTyrValAsnGlnThrSerArgSerAsnValValProValGluAsnPh 277  
1045 ACTGCTATTCCAAACACATCTACAAATCGGCTTAATCTCATGAAACATA 1094  
277 ePheGlnAspAlaTyrLeuGlyValLeuProLysPheSerTyrIleAsn 294  
1095 CATGCTAGAT.....TTTAGTTAT..... 1113  
294 roSerCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 310  
1114 .....AATCCCAATCT..... 313  
311 SerTyrGly..... 313  
1138 TCAAGAGGCTCTCAAAATCGATTCGAATTCATCTTAATATCCCATTTAT 1187  
314 .....GluValPheValLysGlnIleTyrAsp..... 322  
1188 TTCACCTCTATTAGAATTCATTTTATAAGCAATAGCGGATCTCTCTG 1237  
323 .....AlaIleArgGlnGly 327  
1238 TTTTACTTAATTTTAAAGATGGCACTCAAGGTATGCTATGCCCCACCA 1287  
328 ProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGly 344  
1288 CCA.....ACAGGAGG 1298  
344 yPheTyrAspHis.....ValProProLeuAlaValArgProAspA 359  
1299 TGCTGGGACCATCTTTTATGATCTGATGTCGCCACCAAGGCGCAT 1348  
359 snLeuThrTyrThrGluThrAlaLysAsnGly..... 369  
1349 AATTAACATATTTATCTCTCCAGGTGATACATTAAGAGATTTTCATC 1398  
370 GlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 386  
1399 AATGTATATCTCTTATTAAGTACTCCAACTATAAATGAATCACTACACA 1448  
386 pValIleSerProTyr...SerLysLysGlyTyrIleGluGlnTyrGly 402  
1449 AAAAATCAAGGCTTCTCGCGGAAAAGGATATATCAAAATCAAGGGA 1498  
402 hrAspProValThrGlyLysPro 409  
1499 TCATGAATATTTACGGTARCCA 1521

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-158-232-5

seq\_documentation\_block:

: Sequence 5, Application US/08158232  
: Patent No. 5596071  
: GENERAL INFORMATION:  
: APPLICANT: Payne, Jewel  
: APPLICANT: Kennedy, M. Keith  
: APPLICANT: Randall, John Brooks  
: APPLICANT: Meier, Henry  
: APPLICANT: Uick, Heidi Jane  
: APPLICANT: Foncecrada, Luis  
: APPLICANT: Schnopf, H. Ernest  
: APPLICANT: Schwab, George E.  
: APPLICANT: Fu, Jenny  
: TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active  
: TITLE OF INVENTION: Against Hymenopteran Pests  
: NUMBER OF SEQUENCES: 51  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: David R. Saliwanchik  
: STREET: 2421 N.W. 41st Street, Suite A-1  
: CITY: Gainesville  
: STATE: FL  
: COUNTRY: USA  
: ZIP: 32606  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08158,232  
: FILING DATE:  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/887,980  
: FILING DATE: 22-MAY-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/797,645  
: FILING DATE: 25-NOV-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/703,977  
: FILING DATE: 22-MAY-1991  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Saliwanchik, David R.  
: REGISTRATION NUMBER: 31,794  
: REFERENCE/DOCKET NUMBER: M/SCJ104.C1  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 904-375-8100  
: TELEFAX: 904-372-5800  
: INFORMATION FOR SEQ ID NO: 5:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3771 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)  
: HYPOTHETICAL: NO  
: ANTI-SENSE: NO  
: ORIGINAL SOURCE:  
: ORGANISM: Bacillus thuringiensis  
: INDIVIDUAL ISOLATE: 33F2  
: IMMEDIATE SOURCE:  
: CLONE: E. coli NM522(pMYC2316) B-18785  
: FEATURE:  
: NAME/KEY: misc\_feature  
: LOCATION: 4..24  
: OTHER INFORMATION: /function= "oligonucleotide"  
: OTHER INFORMATION: hybridization probe  
: OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
: OTHER INFORMATION: /standard\_name= "probe a"  
: OTHER INFORMATION: /note= "Probe A"  
: FEATURE:  
: NAME/KEY: misc\_feature  
: LOCATION: 13..33  
: OTHER INFORMATION: /function= "oligonucleotide"

```

; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"
; OTHER INFORMATION: /standard_name= "Probe B"
; OTHER INFORMATION: /label= probe-b
; OTHER INFORMATION: /note= "probe b"
US-08-158-232-5

```

## alignment\_scores:

```

Quality: 107.00      Length: 491
Ratio: 0.493         Gaps: 22
Percent Similarity: 44.196      Percent Identity: 19.552

```

## alignment\_block:

```
US-09-426-072-2 x US-08-158-232-5
```

```
Align seg 1/1 to: US-08-158-232-5 from: 1 'to: 3771
```

```

6 LeuLeuThrGlyLeuGlyLeuLeuAlaSerLeuGly...LeuAlaSerPr 21
220 GTTTATCTGTATTAACTTTAATCTCTGAAGTTGGTACTCTGGCTCTGC 269
21 oValThrSerGluThrThrSerValArgGluAlaProPheGlyTyrLysp 38
270 AGCAGTACTATTGTAAGTTTATTGGCCCTAAATATTGGAGATAAAC 319
38 roGlySerLysGluSerIleGluLeuLeuLysAspLysValGluAsnIle 54
320 CAATGCGCAAAATATATTGAAGAGCTCAAGCTCAAAATTCAGCATTA 369
55 ValTrpLeuIleLeuGluAsn...ArgSerPh 64
370 ATTCAACAACATATACAACTATCAAGATGCAATTAATCAAAAATTT 419
64 eAspAsnIleLeuGlyValArgGlnGlyLeuAspAsnProIleA 81
420 TGACAGTCTCAGAAACAATTAATCTATATACAGTA.....GCTATAG 463
81 snAsnGlyProPheCysAsnTyrLysAlaSerAspProSerSerGly 97
464 ATAACAATGATTACGTACACAGCAAA..... 489
98 LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAsnAsp... 112
490 .....ACGCAACTCGAAATCTAAATCTATCTACTTACCTCAGATAT 530
113 .....ProAspHisSerValThrGlyAsnLeuLeuGluPheT 125
531 CTCCATATTTATTCAGAGAGGATAGAACTGGA...GTTTACCTTATT 577
125 YrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141
578 ATGCTATGGTTGCTATGCTCATATATTATTGTTAAGAGACGCTATAGTT 627
142 ....AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLy 157
628 AATCGACAGAAATTAGGCTTTAGTGATAAAGTAGACACACATAAAAA 677
157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGlu 174
678 ATAT.....ATCAAAATGACATACACAAATCATCATCTAGAG 712
174 LuGluValProThrLeuValAspLeuValAspGluPhe..... 186
713 CAGTAATAAAGCATTTCTTAATGGACTTGACAATTTAAGAGCTTTAGAT 762
187 ...ThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn... 201
763 GTAATAGCTATATAAACAACAAATTTAATTAAGGTATGACAGAAAT 812
202 .ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217
813 GGTCTTGATCTAGTTGCTCTATGCGCAACTTTTCGATCCAGATCATTC 862

```

```

217 ysAsnAspAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePhe 233
863 AAAAAGAAAGTAGAAATTTGAATTTACAAGAACTATTCTCTCCAATTTAC 912
234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAs 250
913 CAACCTGTACTTAA.....AACATGCAAAATACCTC 944
250 nGlyGluPheGluProAspSerLeuPhe..... 259
945 TAGCTCTATTGTACTAGCGATCTATTTCACATATCAAGGAGATCTGTAA 994
260 .....Phe 260
995 AATTAGAATTTTCTACAGAACGACACAGATGCTCTTGCAAAATTTT 1044
261 ThrTyrValAsnGlnThrSerArgSerAsnValValProValGluAsnPh 277
1045 ACTGTATTGCAACACATTTACAAATCGCTAATACTATCATGAAACATA 1094
277 ePheGlnAspAlaTyrLeuGlyValLeuProLysPheSerTyrIleAsn 294
1095 CCATGTAGAT.....TTTAGTTAT..... 1113
294 roSerCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 310
1114 .....AATACCAATCT.....AGTGGTAATATT 1137
311 SerTyrGly..... 313
1138 TCAGAGGGCTCTCAAAATCCGATTCGAATTCATTTAATATCCCATTTAT 1187
314 .....GluValPheValLysGlnIleTyrAsp..... 322
1188 TTCAACTTGATTTAGAATTTCAATTTATAAGCAATAGCGGATCTCTG 1237
323 .....AlaIleArgGlnGly 327
1238 TTTTATTAAATTTAAAGATGCGCACTCAAGGCTATGCAATTCGCCAAGCA 1287
328 ProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGly 344
1288 CCA.....ACAGGAGG 1298
344 yPheTyrAspHis.....ValProProProLeuAlaValArgProAsp 359
1299 TCCCTGGGACCAATCTTTTATTGAATCTGATGTCGCCCAAGAGGGCATA 1348
359 snLeuThrTyrThrGluThrAlaLysAsnGly..... 369
1349 AATTAACTATATTTATCTCTCCAGGTGATACATTAAAGAGATTTTCATC 1398
370 GlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 386
1399 AATGTATATCTCTTAAAGTACTCTCAACTATAAATGAAGTATCAACACA 1448
386 pValIleSerProTyr...SerLysLysGlyTyrIleGluGlnTyrGly 402
1449 AAAAATCAAGGCTTCTCGCGAAAGAGATATATCAAAATCAAGGGA 1498
402 hrAspProValThrGlyLysPro 409
1499 TCATCAAAATATTACGGTAAACCA 1521

```

```

seq_name: /cgn2_5/ptodata/2/ina/5A_comb.seq:US-08-304-626-5
seq_documentation_block:

```

```

; Sequence 5, Application US/08304626
; Patent No. 5616495
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Kennedy, M. Keith

```



APPLICANT: Randall, John Brooks  
 APPLICANT: Meier, Henry  
 APPLICANT: Ulick, Heidi Jane  
 APPLICANT: Foncerrada, Luis  
 APPLICANT: Schneff, Harry E.  
 APPLICANT: Schwab, George E.  
 TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates  
 TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding  
 TITLE OF INVENTION: Hymenopteran-Active Toxins  
 NUMBER OF SEQUENCES: 39  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: David R. Saliwanchik  
 STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 32606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/304,626  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/887,980  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Saliwanchik, David R.  
 REGISTRATION NUMBER: 31,794  
 REFERENCE/DOCKET NUMBER: M/SCJ 104  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 904-375-8100  
 TELEFAX: 904-372-5800  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3771 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Bacillus thuringiensis  
 INDIVIDUAL ISOLATE: 33F2  
 IMMEDIATE SOURCE:  
 CLONE: E. coli NM522(pMYC2316) B-18785  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 4..24  
 OTHER INFORMATION: /function= "oligonucleotide  
 OTHER INFORMATION: hybridization probe"  
 OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
 OTHER INFORMATION: /standard\_name= "probe a"  
 OTHER INFORMATION: /note= "probe A"  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 13..33  
 OTHER INFORMATION: /function= "oligonucleotide  
 OTHER INFORMATION: hybridization probe"  
 OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
 OTHER INFORMATION: /standard\_name= "probe B"  
 OTHER INFORMATION: /label= "probe b"  
 OTHER INFORMATION: /note= "probe B"  
 US-08-304-626-5

alignment\_scores: Quality: 107.00 Length: 491  
 Ratio: 0.493 Caps: 22

Percent Similarity: 44.196 Percent Identity: 19.552  
 alignment\_block:  
 US-09-426-072-2 x US-08-304-626-5 ..  
 Align seg 1/1 to: US-08-304-626-5 from: 1 to: 3771  
 6 LeuLeuThrGlyLeuGlyLeuLeuAlaSerLeuGly...LeuAlaSerPr 21  
 220 GTTTTCTGTATTAACCTTTAATCTCGAAGTTGGTACTGGCTCTGC 269  
 21 oValThrSerGluTyrThrSerValArgGluAlaPropheGlyTyrLysP 38  
 270 AGCAAGTACTATTGTAAGTTTATTTGGCCCTAAAAATATTGGAGATAAC 319  
 38 roGlySerLysGluSerIleGluLeuLeuLysAspLysValGluAsnIle 54  
 320 CAATGCAAAAAATATATTTTGAAGAGCTCAAGCTCAAAATTCAGCATTA 369  
 55 ValTrpLeuLeuLeuGluAsn.....ArgSerPh 64  
 370 ATTCAACAAGATATAAACAACCTATCAAGATGCAATTAATCAAAAAAATT 419  
 64 eAspAsnIleLeuGlyGlyValArgGlnGlyLeuAspAsnProIleA 81  
 420 TGACAGTCTTCAGAAACAATTAATCTATATACAGTA.....GCTATAG 463  
 81 snAsnGlyProPheCysAsnTyrLysAsnAlaSerAspProSerSerGly 97  
 464 ATAACAATGATTACGTAAACAGCAAA..... 489  
 98 LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAsnAsp..... 112  
 490 .....ACGCAACTCGAAATCTAAATCTATCTTACCTCAGATAT 530  
 113 .....ProAspHisSerValThrGlyAsnAsnLeuGluPheT 125  
 531 CTCATATATTTATCCAGAAGGATATGAACCTGGA...GGTTTACCTTATT 577  
 125 yrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141  
 578 ATGCTATGTTGCTAATGCTCATATATATTATTGTTAAGAGACGCTATAGTT 627  
 142 ...AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProly 157  
 628 AATGCAGAGAATTAGGCTTTAGTATAAAGAGTAGACACACATATAAAA 677  
 157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGluG 174  
 678 ATAT.....ATCAAAATGACAATACACAATCATACTGAAG 712  
 174 LuGluValProThrLeuValAspLeuValAspGluPhe..... 186  
 713 CAGTAATAAAGCATTCCTTAATGCGACTTCACAAATTTAAGAGTTTATAGAT 762  
 187 ...ThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn... 201  
 763 GTAATAGCTATATAATAAAGCAAAATTTATTTAAAGGTATGACAGAAAT 812  
 202 ..ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217  
 813 GGTTCCTTGATCTAGTTGCTCTATGCGCAACTTTTCGATCCAGATCATATC 862  
 217 ysAsnAspAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePhe 233  
 863 AAAAAAGAGTAGAAATTCGAATTTACAGAACTATTTCTTCTCCAAATTTAC 912  
 234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAs 250  
 913 CCACTGTACCTAAA.....ACATGCAAAATAACCTC 944  
 250 nGlyGluPheGluProAspSerLeuPhe..... 259

```
945 TAGCTCTATTGTACCTAGCGATCTATTTCATCATCAAGGAGATCTGTAA 994
260 .....
995 AATTAGAAATTTTACAAGACGACACGATGGCTTGCACAAATTTT 260
|||
261 ThrTyValAsnGlnThrSerArgSerAsnValProValGluAsnPh 1044
||| .....
1045 ACTGTATTGGAACACATTTTCAAAATCGCTTAATCTCATGAAACATA 1094
277 ePheGlnAspAlaTyLeuGlyValLeuProLysPheSerTyIleAsn 294
|||
1095 CCATGTAGAT .....TTAGTTAT ..... 1113
294 roSerCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 310
||| .....
1114 .....ATACCCAATCT .....AGTGTATAT 1137
311 SerTyGly .....
1138 TCAGAGGCTTCCAAATCGATTCCAAATTCATTTAATCCCAT 313
314 .....GluValPheValLysGlnIleTyAsp ..... 322
1188 TTCAACTTGTATTAGAAATTCATTTATAAGCAATAGCGGATCTCTG 1237
323 .....
1238 TTTTAGTATTATTAAGATGCACCTCAAGGATGTCATTTGCCCAAGCA 1287
328 ProGlnTrpAspLysThrLeuPheIleThrTyAspGluThrGly 344
|||
1288 CCA .....ACAGGAG 1298
344 yPheTyAspHis .....ValProProLeuAlaValArgProAsp 359
1299 TGCTGGGACCATCTTTTATGAATCTGATGTCGCCCAAGGCGCAT 1348
359 snLeuThrTyThrGluThrAlaLysAsnGly ..... 369
1349 AATAAATCTATTTTACTCTCCAGGTATACATTAAGATTCATC 1398
370 GlnLysTyThrLeuHisPheAspArgLeuGlyArgMetProThr 386
||| .....
1399 AATGTATATCTTTATAGTACTCCAACTATAAATGAATCAACAGA 1448
386 pValIleSerProTy .....SerLysLysGlyTyIleGluGlnTyGly 402
||| .....
1449 AAAATCAAGGCTTCTCGGAAAAAGGATATATCAAAAATCAAGGA 1498
402 hrAspProValThrGlyLysPro 409
1499 TCATGAATATTTACGCTAACCA 1521
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-316-301A-5
seq_documentation_block:
; Sequence 5, Application US/08316301A
; Patent No. 5753492
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schwab, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Poncetrada, Luis
; TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes
; NUMBER OF INVENTION: Which Code Therefor
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
```

```
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,301A
FILING DATE: 30-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/871,510
FILING DATE: 23-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/693,018
FILING DATE: 03-MAY-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/565,544
FILING DATE: 10-AUG-1990
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/084,653
FILING DATE: 12-AUG-1987
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/830,050
FILING DATE: 31-JAN-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lloyd, Jeff
REGISTRATION NUMBER: 35,589
REFERENCE/DOCKET NUMBER: MA20CCCD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 5 (PS33F2):
SEQUENCE CHARACTERISTICS:
LENGTH: 3771 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
INDIVIDUAL ISOLATE: 33f2
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC 2316) B-18785
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4..24
OTHER INFORMATION: /function= oligonucleotide
OTHER INFORMATION: hybridization probe"
OTHER INFORMATION: /product= GCA/T ACA/T TTA AAT GAA GTA/T TAT"
OTHER INFORMATION: /standard_name= probe a"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13..33
OTHER INFORMATION: /function= oligonucleotide
OTHER INFORMATION: hybridization probe"
OTHER INFORMATION: /product= AAT GAA GTA/T TAT CCA/T GTA/T AAT"
OTHER INFORMATION: /standard_name= probe b"
OTHER INFORMATION: /label= probe-b
US-08-316-301A-5
alignment_scores:
```



1 NUMBER OF SEQUENCES: 51  
 2 CORRESPONDENCE ADDRESS:  
 3 ADDRESSEE: David R. Saliwanchik  
 4 STREET: 2421 N.W. 41st Street, Suite A-1  
 5 CITY: Gainesville  
 6 STATE: FL  
 7 COUNTRY: USA  
 8 ZIP: 32606  
 9 COMPUTER READABLE FORM:  
 10 MEDIUM TYPE: Floppy disk  
 11 COMPUTER: IBM PC compatible  
 12 OPERATING SYSTEM: PC-DOS/MS-DOS  
 13 SOFTWARE: Patent In Release #1.0, Version #1.25  
 14 CURRENT APPLICATION DATA:  
 15 APPLICATION NUMBER: US/08/611,928  
 16 FILING DATE: 06-MAR-1996  
 17 CLASSIFICATION: 530  
 18 PRIOR APPLICATION DATA:  
 19 APPLICATION NUMBER: US 08/158,232  
 20 FILING DATE: 24-NOV-1993  
 21 APPLICATION NUMBER: US 07/887,980  
 22 FILING DATE: 22-MAY-1992  
 23 CLASSIFICATION: 530  
 24 PRIOR APPLICATION DATA:  
 25 APPLICATION NUMBER: US 07/797,645  
 26 FILING DATE: 25-NOV-1991  
 27 CLASSIFICATION: 530  
 28 PRIOR APPLICATION DATA:  
 29 APPLICATION NUMBER: US 07/703,977  
 30 FILING DATE: 22-MAY-1991  
 31 CLASSIFICATION: 530  
 32 ATTORNEY/AGENT INFORMATION:  
 33 NAME: Saliwanchik, David R.  
 34 REGISTRATION NUMBER: 31,794  
 35 REFERENCE/DOCKET NUMBER: M/SCI104.C1  
 36 TELECOMMUNICATION INFORMATION:  
 37 TELEPHONE: 904-375-8100  
 38 TELEFAX: 904-372-5800  
 39 INFORMATION FOR SEQ ID NO: 5:  
 40 SEQUENCE CHARACTERISTICS:  
 41 LENGTH: 3771 base pairs  
 42 TYPE: nucleic acid  
 43 STRANDEDNESS: double  
 44 TOPOLOGY: linear  
 45 MOLECULE TYPE: DNA (genomic)  
 46 HYPOTHETICAL: NO  
 47 ANTI-SENSE: NO  
 48 ORIGINAL SOURCE:  
 49 ORGANISM: Bacillus thuringiensis  
 50 INDIVIDUAL ISOLATE: 33F2  
 51 CLONE: E. coli NM522(pMYC2316) B-18785  
 52 FEATURE:  
 53 NAME/KEY: misc.feature  
 54 LOCATION: 4..24  
 55 OTHER INFORMATION: /function= "oligonucleotide"  
 56 OTHER INFORMATION: /hybridization probe  
 57 OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
 58 OTHER INFORMATION: /standard\_name= "probe a"  
 59 OTHER INFORMATION: /note= "Probe A"  
 60 FEATURE:  
 61 NAME/KEY: misc.feature  
 62 LOCATION: 13..33  
 63 OTHER INFORMATION: /function= "oligonucleotide"  
 64 OTHER INFORMATION: /hybridization probe  
 65 OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
 66 OTHER INFORMATION: /standard\_name= "Probe B"  
 67 OTHER INFORMATION: /label= "probe-b"  
 68 OTHER INFORMATION: /note= "probe b"  
 69 US-08-611-928-5

alignment\_scores:

Quality: 107.00 Length: 491  
 Ratio: 0.493 Gaps: 22  
 Percent Similarity: 44.196 Percent Identity: 19.552  
 alignment\_block:  
 US-09-426-072-2 x US-08-611-928-5  
 Align seg 1/1 to: US-08-611-928-5 from: 1 to: 3771  
 6 LeuLeuThrGlyLeuGlyLeuLeuAlaSerLeuGly...LeuAlaSerPr 21  
 220 GTTTTATCTGTATTAACTTTAATCTCTGAAGTTGGTACTGTGCCTCTGC 269  
 21 oValThrSerGluTyrThrSerValArgGluAlaProPheGlyTyrLysp 38  
 270 AGCAAGTACTATTGTAAGTTTATTGGCTTAAATATTTGGAGATAAAC 319  
 38 roGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIle 54  
 320 CAATGCCAAATAATATTTGAAGCTCAAGCCTCAATGAAGCATTA 369  
 55 ValTrpLeuIleLeuGluAsn.....ArgSerPh 64  
 370 ATTCAACAAGATATAACAACTATCAAGATGCAATTAATCAAAAAAATT 419  
 64 easpAsnIleLeuGlyGlyValArgGlnGlyLeuAspAsnProIleA 81  
 420 TGACAGTCTTCAGAAACAAATTAATCTATATACAGTA.....GCTATAG 463  
 81 snAsnGlyProPheCysAsnTyrLysAsnAlaSerAspProSerSergly 97  
 464 ATAACAATGATTACGTAACAGCAAAA..... 489  
 98 LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAsnAsp..... 112  
 490 .....ACGCAACTCGAAATCTAAATCTTACTTACCTCAGATAT 530  
 113 .....ProAspHisSerValThrGlyAsnAsnLeuGluPheT 125  
 531 CTCCTATATTATCCAGAAGGATATGAACCTGGA...GGTTTACCTTATT 577  
 125 YrGlyThrTyrTrpProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141  
 578 ATGCTATGGTGGCTAATGCTCATATATATTGTTAAGACAGCCTATAGTT 627  
 142 ...AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLy 157  
 628 AATCCAGAGAAATTAGCTTTAGTGATAAAGAAGTAGACACACATAAAAA 677  
 157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGlu 174  
 678 ATAT.....ATCAAAATGACAATACACAATCATCTAG 712  
 174 LuGluValProThrLeuValAspLeuValAspGluPhe..... 186  
 713 CAGTAATAAAGCATCTTAAATGAGCTTGACAAATTTAAGAGTTTAGAT\*762  
 187 ...ThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn.. 201  
 763 GTAAATAGCTATATAAATAAAGCAAAATATATATAAAGGTATGACAGAAAT 812  
 202 ..ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217  
 813 GGTTCITGATCTAGTTGCTCTATGGCCAACCTTCGATCCAGATCATATTAT 862  
 217 ysAsnAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePhe 233  
 863 AAAAAGAGTAGAAATTAATTTACAAGAACATTTCTTCTCCCAATTTAC 912  
 234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAs 250  
 913 CAACCTGTACCTAAA.....AACATGCAAAATACCTC 944



## alignment\_block:

US-09-426-072-2 x US-09-173-891-5

Align seg 1/1 to: US-09-173-891-5 from: 1 to: 3771

```

6 LeuLeuThrGlyLeuGlyLeuAlaSerLeuGly...LeuAlaSerPr 21
  ::::::::::: ||| ::::::::::: :::::::::::
220 GTTATCTGTAATTAACCTTAATTCCTGAAGTTGGTACTGGCCCTCG 269
  ::::::::::: ||| ::::::::::: :::::::::::
21 oValThrSerGlyThrSerValArgGluAlaProPheGlyTyrLysp 38
  ::::::::::: ||| ::::::::::: :::::::::::
270 AGCAAGTACTATTGTAAGTTTATTTGGCTAAATATTGGAGATAAAC 319
  ::::::::::: ||| ::::::::::: :::::::::::
38 roGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIle 54
  ::::::::::: ||| ::::::::::: :::::::::::
320 CAATGCGAANAATATATTGAGAGCTCAGCTCAATGAATGAAGCATTA 369
  ::::::::::: ||| ::::::::::: :::::::::::
55 ValTrpLeuIleLeuGluAsn.....ArgSerPh 64
  ::::::::::: ||| ::::::::::: :::::::::::
370 ATTCAACAAGATATAACAACACTATCAAGATGCAATTAATCAAAAAAAT 419
  ::::::::::: ||| ::::::::::: :::::::::::
64 eAspAsnIleLeuGlyGlyValArgGlnGluLeuAspAsnProIleA 81
  ::::::::::: ||| ::::::::::: :::::::::::
420 TGACAGTCTTCAGAAACAATTAATCTATATACAGTA.....GCTATAG 463
  ::::::::::: ||| ::::::::::: :::::::::::
81 snAsnGlyProPheCysAsnTyrLysAsnAlaSerAspProSerSerGly 97
  ::::::::::: ||| ::::::::::: :::::::::::
464 ATACCAATGATTACGTACACGCAAAA..... 489
  ::::::::::: ||| ::::::::::: :::::::::::
98 LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAsnAsp..... 112
  ::::::::::: ||| ::::::::::: :::::::::::
490 .....ACGCACTCGAANAATCTAATTTCTATCTTACCTCAGATAT 530
  ::::::::::: ||| ::::::::::: :::::::::::
113 .....ProAspHisSerValThrGlyAsnAsnLeuGluPheT 125
  ::::::::::: ||| ::::::::::: :::::::::::
531 CTCATATATTATCCAGAGGATGAAACTGGA...GGTTTACCTTATT 577
  ::::::::::: ||| ::::::::::: :::::::::::
125 YrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141
  ::::::::::: ||| ::::::::::: :::::::::::
578 ATGCTATGTTGCTAATGCTCATATATTATTGTTAAGAGACGCTATAGT 627
  ::::::::::: ||| ::::::::::: :::::::::::
142 ...AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProly 157
  ::::::::::: ||| ::::::::::: :::::::::::
628 AATGACAGAAATTAGCTTTAGTGTAAAGAGTAGACACACATAAAAA 677
  ::::::::::: ||| ::::::::::: :::::::::::
157 sLeuAlaProGluAlaThrArgGlnValMetGlyTyrTyrThrGluG 174
  ::::::::::: ||| ::::::::::: :::::::::::
678 ATAT.....ATCAAAATGACAATACACAATCATCATCTGAAG 712
  ::::::::::: ||| ::::::::::: :::::::::::
174 LuGluValProThrLeuValAspLeuValaspGluPhe..... 186
  ::::::::::: ||| ::::::::::: :::::::::::
713 CAGTAATAAAGCATCTTAATGAGCTTGCAAAATTTAAGAGTTTAGAT 762
  ::::::::::: ||| ::::::::::: :::::::::::
187 ...ThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn... 201
  ::::::::::: ||| ::::::::::: :::::::::::
763 GTAATAGCTATAATAAAGCAAAATATATTAAAGGTATGACAGAAAT 812
  ::::::::::: ||| ::::::::::: :::::::::::
202 .ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217
  ::::::::::: ||| ::::::::::: :::::::::::
813 GGTCTTCTGATCTAGTTGCTATGCCCACTTTCGATCCAGATCATATATC 862
  ::::::::::: ||| ::::::::::: :::::::::::
217 yAsnAspAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePhe 233
  ::::::::::: ||| ::::::::::: :::::::::::
863 AAAAGAAGTAGAATTTGAATTTACAAGAACTATTCTTCTCCAATTTAC 912
  ::::::::::: ||| ::::::::::: :::::::::::
234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAs 250
  ::::::::::: ||| ::::::::::: :::::::::::
913 CAACCTGTACCTAAA.....ACATGCAAAATACCTC 944
  ::::::::::: ||| ::::::::::: :::::::::::
250 nGlyGluPheGluProAspSerLeuPhe..... 259
  ::::::::::: ||| ::::::::::: :::::::::::
945 TACCTCTATTGTACCTAGCGATCTATTTCACATATCAAGGAGATCTTTGTA 994
  ::::::::::: ||| ::::::::::: :::::::::::

```

```

260 .....Phe 260
995 AATTAGAAATTTTCTACAGAACGACACAGATGCTTGCAAAAATTTT 1044
  ::::::::::: ||| :::::::::::
261 ThrTyrValAsnGlnThrSerArgSerAsnValValProValGluAsnPh 277
  ::::::::::: ||| :::::::::::
1045 ACTGGTATTCGAAACACATCTTACAAATCGCCTAATCTCATGAAACATA 1094
  ::::::::::: ||| :::::::::::
277 ePheGlnAspAlaTyrLeuGlyValLeuProLysPheSerTyrIleAsnP 294
  ::::::::::: ||| :::::::::::
1095 CCATGTAGAT.....TTTAGTTAT..... 1113
  ::::::::::: ||| :::::::::::
294 roSerCysCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 310
  ::::::::::: ||| :::::::::::
1114 .....AATACCAATCT.....AGTGGTATATAT 1137
  ::::::::::: ||| :::::::::::
311 SerTyrGly..... 313
  ::::::::::: ||| :::::::::::
1138 TCAAGAGGCTCTTCAAATCCGATTCCAATTCATCTTAATAATCCCATTTAT 1187
  ::::::::::: ||| :::::::::::
314 .....GluValPheValLysGlnIleTyrAsp..... 322
  ::::::::::: ||| :::::::::::
1188 TTCAACTTCTATTAGAAAATTCATTTATAAGGCAATAGCGGGATCTCTG 1237
  ::::::::::: ||| :::::::::::
323 .....AlaIleArgGlnGly 327
  ::::::::::: ||| :::::::::::
1238 TTTTAGTTAATTTTAAAGATGGCACTCAAGGGTATGCAATTTGCCCAAGCA 1287
  ::::::::::: ||| :::::::::::
328 ProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyG 344
  ::::::::::: ||| :::::::::::
1288 CCA.....ACAGGAGG 1298
  ::::::::::: ||| :::::::::::
344 yPheTyrAspHis.....ValProProProLeuAlaValArgProAspA 359
  ::::::::::: ||| :::::::::::
1299 TGCTTGGGACCATCTTTTATTGAATCTGATGTCGCCCAAGAGGCATA 1348
  ::::::::::: ||| :::::::::::
359 snLeuThrTyrThrGluThrAlaLysAsnGly..... 369
  ::::::::::: ||| :::::::::::
1349 AATTAAACTATATTTTATCTCTCCAGGTGATACATTAAAGAGATTTTCATC 1398
  ::::::::::: ||| :::::::::::
370 GlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 386
  ::::::::::: ||| :::::::::::
1399 AATGTATATATCTTATAGTACTCCCACTATAAATGAACATATCAACAGA 1448
  ::::::::::: ||| :::::::::::
386 pValIleSerProTyr...SerLysLysGlyTyrIleGluGlnTyrGlyT 402
  ::::::::::: ||| :::::::::::
1449 AAAAATCAAGGCTTTCTCGGGAAGAGGATATATCAAAAATCAAGGGA 1498
  ::::::::::: ||| :::::::::::
402 hrAspProValThrGlyLysPro 409
  ::::::::::: ||| :::::::::::
1499 TCATGAAATATTACGGTAAACCA 1521
  ::::::::::: ||| :::::::::::

```

seq\_name: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:US-09-076-137-5

## seq\_documentation\_block:

```

; Sequence 5, Application US/09076137B
; Patent No. 6166195
; GENERAL INFORMATION:
; APPLICANT: Schnef, Harry E.
; APPLICANT: Schwab, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Fongerrada, Luis
; TITLE OF INVENTION: No. 6166195el Nematode-Active Toxins and Genes Which Code
; FILE REFERENCE: MA-20CCCD2
; CURRENT APPLICATION NUMBER: US/09/076.137B
; EARLIER FILING DATE: 1998-05-12
; EARLIER APPLICATION NUMBER: 08/316,301
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1

```



```

402 hrAspProValThrGlyLysPro 409
|||||
1499 TCATGAAATATTACGGTAACCA 1521

seq_name: /cgn2_6/ptodata/2/lna/PCITUS_COMB.seq:PCT-US92-03624-5
seq_documentation_block:
; Sequence 5, Application PC/TUS9203624
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schnepf, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Focerrada, Luis
; TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes
; TITLE OF INVENTION: Which Code Therefor
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/03624
; FILING DATE: 19920501
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA20C2C1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 5 (PS33F2):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3771 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: 33f2
; IMMEDIATE SOURCE:
; CLONE: E. COLI NM522(pMYC 2316) B-18785
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4..24
; OTHER INFORMATION: /function= "oligonucleotide"
; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"
; OTHER INFORMATION: /standard_name= "probe a"
; OTHER INFORMATION: /note= "Probe A"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13..33
; OTHER INFORMATION: /function= "oligonucleotide"
; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"
; OTHER INFORMATION: /standard_name= "Probe B"
; OTHER INFORMATION: /label= "probe-b"
; OTHER INFORMATION: /note= "probe b"
PCT-US92-03624-5

```

```

alignment_scores:
  Quality: 107.00      Length: 491
  Ratio: 0.493        Gaps: 22
  Percent Similarity: 44.196      Percent Identity: 19.552

alignment_block:
US-09-426-072-2 x PCT-US92-03624-5 ..
Align seg 1/1 to: PCT-US92-03624-5 from: 1 to: 3771

6 LeuLeuThrGlyLeuGlyLeuLeuAlaSerLeuGly...LeuAlaSerPr 21
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
220 GTTTTATCTGATTAACTTTAACTTCTGAAGTGTGCTACTGTGCTCTGC 269
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

21 oValThrSerGluTyrThrSerValArgGluAlaProPheGlyTyrLysp 38
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
270 ACCAGTACTATTCTTAAGTTTATTGGCTTAAATATTGGAGATAAAC 319
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

38 roGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIle 54
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
320 CAAATGCAAAATATATTGAAGAGCTCAAGCTCAAAATTGAAGCATT 369
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

55 ValTrpLeuIleLeuGluAsn.....ArgSerPh 64
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
370 ATTCAACAAGATATAACAACACTATCAAGATGCAATTAATCAAAAAAATT 419
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

64 eAspAsnIleLeuGlyValArgGlnGlyLeuAspAsnProIleA 81
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
420 TCACAGCTCTTCAGAAACAAATTAATCTATATACAGTA.....GCTATAG 463
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

81 snAsnGlyProPheCysAsnTyrLysAsnAlaSerAspProSerSerCly 97
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
464 ATAACAATGATTACGTAACACAAAA..... 489
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

98 LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAsnAsp..... 112
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
490 .....ACGCAACTCGAAATCTAAATCTATATCTACTTACCTCAGATAT 530
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

113 .....ProAspHisSerValThrGlyAsnAsnLeuGluPheT 125
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
531 CTCCATATTATTCCAGAGGATATGAAGCTGCA...GGTTACCTATT 577
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

125 YGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
578 ATGCTATGGTTCATATGCTCATATATTATTGTAAGAGACGCTATAGTT 627
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

142 ...AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLy 157
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
628 AATGCAGAGAAATTAGGCTTTAGTGATAAAGAAAGTAGACACACATAAAA 677
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGlu 174
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
678 ATAT.....ATCAAAATGACATACACAAATCATACTGAG 712
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

174 LuGluValProThrLeuValAspLeuValAspGluPhe..... 186
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
713 CAGTAATAAAGCACTCTTAAATGGAGCTTGACAAATTTAAGAGTTAGAT 762
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

187 ...ThrThrPheAsnSerTyrPheSerCysValProGlyProThrAsn... 201
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
763 GTAAATAGCTATATAATAAAGCAATATATATAAGGTATGACAGAAAT 812
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

202 .ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
813 GGTTCTTGATCTAGTTGCTCTATGCGCAACTTTCCATCCAGATCAATTATC 862
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

217 ysAsnAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePhe 233
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||
863 AAAAGAAGTAGAAATTGAATTTACAGAACTATTCTCTCCCAATTTC 912
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAs 250
   :::::  |||  :::::  :::::  |||  :::::  |||  :::::  |||

```



```

913 CAACCTGTACCTAA.....AACATGCAAAATACCTC 944
250 nGlyClupheGluProAspSerLeuPhe..... 259
945 TACCTCTATTGTACCTAGCGATCTATTTCATCATCAAGGAGATCTGTAA 994
260 .....Phe 260
995 AATTAGAATTTTCTACAAGACGGACACGATGGTCTTGCAAAAATTTT 1044
261 ThrTyValAsnGlnThrSerArgSerAsnValProValGluAsnPh 277
1045 ACTGGTATTGCAACACATTTCAAAATCGCTTAATCTACTCATGAAACATA 1094
277 ePheGlnAspAlaTyLeuGlyValLeuProLysPheSerTyLeAsnP 294
1095 CCATGTAGAT.....TTTACTTAT..... 1113
294 roSerCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 310
1114 .....AATACCAATCT.....AGTGGTAATATT 1137
311 SerTyGly..... 313
1138 TCAAGAGCTCTTCAAAATCCGATTCCAATTGATCTTTAATAATPCCATTAT 1187
314 .....GluValPheValLysGlnIleTyAsp..... 322
1188 TTCAACTGTATTAGAAATTTTATTAAGGCAATAGCGGATCTCTG 1237
323 .....AlaIleArgGlnGly 327
1238 TTTTATTGTTAAATTTTAAAGATGGCACTCAAGGGTATGCATTGCCCAAGCA 1287
328 ProGlnTrpAspLysThrLeuPheIleThrTyAspGluThrGlyGI 344
1288 CCA.....ACAGGAGG 1298
344 yPheTyAspHis.....ValProProLeuAlaValArgProAspA 359
1299 TCGCTGGGACCATTTCTTTTATTGAATCTGTGTCGCCCAAGAGGCATA 1348
359 snLeuThrTyThrGluThrAlaLysAsnGly..... 369
1349 AATTAACATATATTTATATCTTCCAGTGATACATTAAGAGATTTCATC 1398
370 GlnLysTyThrLeuHisPheAspArgLeuGlyArgMetProThrTr 386
1399 AATGTATATATCTTTTAAAGTACTCCAACATAAATGAATCAACAGA 1448
386 pValIleSerProTy.....SerLysLysGlyTyIleGluGlnTyGlyT 402
1449 AAAAATCAAGGCTTCTTCGCGGAAAAGGATATATCAAAAATCAAGGGA 1498
402 hrAspProValThrGlyLysPro 409
1499 TCATGAATATTACGGTAAACCA 1521
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-119-361-4

```

```

seq_documentation_block:
; Sequence 4, Application US/08119361
; Patent No. 5523390
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis Arginine-specific Proteinase
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201

```

```

; CITY: Boulder
; STATE: CO
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,361
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 21-93
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-499-8080
; TELEFAX: 303-499-8089
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3159 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 949..3159
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1630..3105
; US-08-119-361-4

```

```

alignment_scores:
  Quality: 104.50      Length: 521
  Ratio: 0.488        Gaps: 30
  Percent Similarity: 41.075  Percent Identity: 19.962

```

```

alignment_block:
US-09-426-072-2 x US-08-119-361-4 ..

Align seg 1/1 to: US-08-119-361-4 from: 1 to: 3159

36 TyLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValGI 52
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1615 TACGAGCGGGCGCTTACACACCGGTAGAG.....GAAAAACA 1652
52 uAsnIleValTrpLeuIleLeuGluAsnArgSerPheAspAsnIleLeuG 69
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1653 AAATGGTGGTATGTCATCGTAGCCAAAAAGATATGAG.....G 1693
69 lyGlyValArgGlnGlyLeuAspAsnProIleAsnAsnGlyProPhe 85
   ||:|||||:|||||:|||||:|||||:|||||:|||||:
1694 GAGATATTAA.....GATTC 1710
86 CysAsnTyLysAsnAlaSerAspProSerSerGlyLysTyCysThrGI 102
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1711 GTTATTGGAACCAACCGCT.....CTCCGTCACGA 1745
102 nAlaLysAspTyAspSerValPheAsnAspProAspHisSerValThrG 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1746 GGTGAA.....GTGCAGAGATATTGCTTCTCCCGTTACAG 1783
119 lyAsnLeuGluPheTy.....GlyThrTyThrProAsnAsnGly 133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1784 CTAATGCTATTTCAGCAGTTCTGTTAAGCAAGAAATACGAAAGAGTAAT 1833
134 AlaIleAlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAl 150

```

```

1834 GATTGACCTATGTTCTTTGGTGGCATCAC..... 1866
      ::::: ::::: ::::: :::::
150 aClnLeuAsnAspTyrPro...LysLeuAlaProGluGluAlaThrArgG 166
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1867 .....AAGATATCTCTGCCAAATTAATCTCCGGGATCAATTCGGACC 1909
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
166 lnValMetGly.....TyrTyrThrGlu..... 173
      ||| ||| ::::: ||| ::::: |||
1910 AGGTATATGGACAAATAGTAGTAATCAACCACTACACGAAGTCTTCATC 1959
      ::::: ||| ::::: ||| ::::: |||
174 .....GluGluValProThrLeuValAspLe 182
      ||| ::::: ||| ::::: |||
1960 GGTCGTTCTCATGTGACAGCAAGAGGATCTGAAGACAAATCATCG 2009
      ::::: ||| ::::: ||| ::::: |||
182 uValAspGluPhe.....ThrThrPheAsnSerTyrPheSerC 195
      ::::: ||| ::::: ||| ::::: |||
2010 GACTATTCACTATGAGCGCAATATATACACGGAAGACAAATGGCTCGGT. 2058
      ::::: ||| ::::: ||| ::::: |||
195 ysValProGlyProThrAsnProAsnArgLeuCys.....AlaLeu 208
      ::::: ||| ::::: |||
2059 .....CAGGCTCTTTGTATGCTTCGGCTGAA 2085
      ::::: ||| ::::: |||
209 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 225
      ::::: ||| ::::: ||| ::::: |||
2086 GGAGGCCCATCCGCAGACAATGTTGAAGTGATATCCAGCATGAGAATGT 2135
      ::::: ||| ::::: |||
225 r..... 225
2136 AATCGCCAATCTGTACCNGATGGCTATACCAAGATTATCAATGTT 2185
      ::::: ||| ::::: |||
226 .....GlyIleSerSerIlePheGluAlaAlaAsnGluLys 239
      ::::: ||| ::::: ||| ::::: |||
2186 ATGATCCGGGAGTAACTCTAAACATTTATGATGTTTCAAC...GGA 2232
      ::::: ||| ::::: |||
240 GlyValSerTyrLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAs 256
      ||| ::::: ||| ::::: |||
2233 GGAATCTGTTGTCACACTATACGGGCCACGGTAGCGGAACACAGCTGGGG 2282
      ::::: ||| ::::: |||
256 pSerLeuPhePhe.....ThrTyrValAsnGlnThrSerArgSerAsnV 271
      ::::: ||| ::::: ||| ::::: |||
2283 TAGCTCTCATTGGCCACCCTCATGTGAAGCAGCTTACCACAGCAACC 2332
      ::::: ||| ::::: |||
271 alValProValGluAsnPheGlnAspAla..... 281
      ::::: ||| ::::: |||
2333 AGCTACCG.....TTTATTTCGACGTAGCTTGTGTGAATGGCGAT 2373
      ::::: ||| ::::: |||
282 TyrLeuGlyValLeuProLysPhe..... 289
      ::::: ||| ::::: |||
2374 TTCCTATTCAAGTGCCTGCTTCGCAGAGCCCTGATGGTCGCACAAA 2423
      ::::: ||| ::::: |||
290 .....SerTyrIleAsnP 294
      ::::: ||| ::::: |||
2424 AGATGGTAAGCGCAGAGTACTGTTGCTATCATAGCGTCTACGATCAACC 2473
      ::::: ||| ::::: |||
294 roSer.....Cys 296
      |||
2474 AGTCTGGCGTTCTCCTATCGCGGCAGGATGAGATGAACGAAATTCGT 2523
      ::::: |||
297 CysGlyThrAsnThrAsnSerMetHisProThr...GlyAsnValSerTy 312
      ||| ::::: ||| ::::: |||
2524 TCGGAAAAACCCGCAACAACATCAAGCGTACTTTTCGGTGGTGCACCAT 2573
      ::::: ||| ::::: |||
312 rGlyGluValPheVal.....LysGlnIleTyrAspAlaIleArgG 326
      ::::: ||| ::::: |||
2574 GAACGTATGTTGCTATGTGTGAAAGATATAAAAGGATGGTGAGAAGA 2623
      ::::: ||| ::::: |||
326 lnGlyProGlnTrp.....AspLysThrLeuLeuPheIleThr 338
      ::::: ||| ::::: |||
2624 TGCTCACACATGGACTGTTTCGGCGCCCTCGCTCGCTCGTAC 2673
      ::::: ||| ::::: |||
339 TyrAspGluThrGlyGlyPheTyrAspHisValProProLeuAlaVal 355
      ::::: ||| ::::: |||

```

\* . . . \*

```

2674 CTT.....GTCCCGACCAAAATGCAGGT 2696
355 l...ArgProAspAsnLeuThrTyrThrGluThr..... 365
      ||| ::::: ||| ::::: |||
2697 TACGGCTCCGGCTCAGATTAAATTTGACGGATGCTTCAGTCAACGTATCTT 2746
      ::::: ||| ::::: |||
366 .....AlaLysAsnGlyGlnLysTyr 372
      ::::: ||| ::::: |||
2747 GCGATTATAATGGTGCTATTGCTTACCATTTTCAGCCAATGGAAAGATGTTT 2796
      ::::: ||| ::::: |||
373 .....ThrLeuHisPheAspArgLe 379
      ||| ::::: |||
2797 GGTTCTGCAGTTGTGCAAAATGGAACAGCTACAATCAATCTGACAGGTCT 2846
      ::::: ||| ::::: |||
379 uGlyGlyArgMetPro...ThrTrpValIleSerProTyrSerLysLysG 395
      ||| ::::: ||| ::::: |||
2847 GACAAATGAAAGCACGCTTACCCTTACAGTAGTTGGTTACAAACAAGAGA 2896
      ::::: ||| ::::: |||
395 lYTyrlleGluGlnTyrGlyThrAspProValThrGlyLysProAlaPro 411
      ||| ::::: |||
2897 CGTTTATTAGACCATCAACACTAAT.....GGTAGCCCTAACCC 2937
      ::::: ||| ::::: |||
412 Tyr.....SerAlaThrSerValLeuLysThrLeuG 422
      ||| ::::: |||
2938 TACCAAGCCCGTTTCCAACTTGACAGCTACAAACGAGGTCAGAAAAGTAAC 2987
      ::::: ||| ::::: |||
422 yTyrLeuTrpAsp 426
      |||
2988 GCTCAAGTGGGAT 3000

```



## RESULT 2

T02648

hypothetical protein F12C20.9 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 14-May-1999

C:Accession: T02648

R:Rounsley, S.D.; Ronning, C.M.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; S.

submitted to the EMBL Data Library, August 1998

A:Description: Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.

A:Reference number: 214685

A:Accession: T02648

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-514 &lt;R0&gt;

A:Cross-references: EMBL:AC005168; NID:g3426033; PID:g3426039

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Insertions: 156/2; 280/3

A:Note: F12C20.9

Query Match 17.4%; Score 435.5; DB 2; Length 514;  
Best Local Similarity 29.8%; Pred. No. 3.3e-24;  
Matches 130; Conservative 79; Mismatches 162; Indels 65; Gaps 17;

QY 51 VENIVWLLENRFDNIGVRRGGLDNPIN--NGPFCNVKNASDPSSGKYCTQAKDYDS 108

DB 27 IKTIIVVVVWNRSDHMLGWMKK--LNPEINGVDGSESINPVSDPSRKRKIGSGSH-Y 83

QY 109 VFNDPDHSVYTGNNLEFYGY-----TPNGATASGKVVADQSGFLNAQLNDYPKLAPEE 162

DB 84 VDPDPGHGFAIRQVFGSNDTMDPPPMNGFVQQA--YSEDPSGNMSA----- 130

QY 163 ATRQVNGVYTEEVPTLVLDVDEFTFNSWFCVPGPTNPRLCALAGTAAGKKNDDDF 222

DB 131 ---SVNGEEDPKVPYKSLVSEFAVDFRWFASVPSQPNRKFVHSGTSAGATSNPIS 187

QY 223 LNYGSSKIFEANEGVSNLWYDGTNGEPEPDSLFYVYVQNT-----SRNVVVPVE 275

DB 188 LAKGYQRTIEDNLDDSEF---GIYQINPAVLFYQSLRKLKYVFKFHSYGN----- 238

QY 276 NFQDAYLGLVPKFSYINPSCGT---NTNSMHTGNVSGVEFVKQIYDAIRGPQWDK 332

DB 239 SFKDHAKQKLPATVIEQRYMDTLLEPASDDHSDHYQGGKFIKEYEYELRASQWNE 298

QY 333 TLLFTTYDETGFGYDHPVPL--AVRPDNLTYTETAKNGOKYTLHFDRLGGRMPTWVISP 390

DB 299 TLLIITYDEHGYDFHVPTRVNPSPDGIV-----GDPFLFQFNRLGIRVPTIAVSP 352

QY 391 YSKKGYIEQYGTDPVTKCAP---YSATSVLTKLGLWLDIED--FTPRVAHSFSDHL--- 443

DB 353 WIEKGV-VHGNP---GSPFSEYEHSSIPATVKLFNLSSPFLTKDEWAGTFENILQ 408

QY 444 IGTTLREDAPIALKTP 459

DB 409 IRKEPTDCPETLPEP 424

## RESULT 3

F70662

probable plcC protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: F70662

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,

Rajandream, M.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: F70662

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-508 &lt;COL&gt;

A:Cross-references: GB:283860; GB:AL123456; NID:g3261681; PIDN:CAB06146.1; PID:g17812

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: plcC

Query Match 13.0%; Score 324.5; DB 2; Length 508;  
Best Local Similarity 26.0%; Pred. No. 4.3e-16;  
Matches 140; Conservative 61; Mismatches 203; Indels 135; Gaps 24;

QY 10 LGLASLGLASPVTSYTSVREAPFGYKPGSKSNIENKDKVENIVWLLENRSDNIG 69

DB 7 LAKAAGAGAAVLTDAAPVIEKAYGAPGCS---GHLTD-IEHIVLCLOENRSDHYFG 61

QY 70 GVRRL--QGLDNPINNGPCNVKNASDPSSGKYCTQAKDYDSV-----F 110

DB 62 TISAVDGFDTPT---PLFOQKGNPE-----TQALDPTGITLPIRYINTTGGPVGECV 112

QY 111 NPDHSHVYTGNNLEFYGYTPNNGATASGKVVADQSGFLNAQLNDYPKLAPEEATROVMGY 170

DB 113 NPDHSHVYTGNNLEFYGYTPNNGATASGKVVADQSGFLNAQLNDYPKLAPEEATROVMGY 153

QY 171 YTEEVPTLVLDVDEFTFNSWFCVPGPTNPRLCALAGTAAGHG--KNDDDFLNYGISS 229

DB 154 YARPDPHYLLADDTICQYFSSLLGTPNRLYWSATVNPDCGGGQFQIVAEPAIOP 213

QY 230 KSIF-----EANEKGVSN-----LNYDGTNGEF---EPDSLFYTYVN 264

DB 214 KLTFTWRIMPONLSDAGISWKVYNSKLLGLNDTSLSRNGYVGSFKAADPRS----- 266

QY 265 QTSRSNVVPE--NFTQDAYLGLVPKFSYINPSCGTNTNSMHTGNVSGVEFVKQIYD 322

DB 267 DLARYGIAPYDWDIFRDVNTLPOYSWVVP---LTVESEHSPFPVAVGAVTIINLR 322

QY 323 ATRQVNGVYTEEVPTLVLDVDEFTFNSWFCVPGPTNPRLCALAGTAAGHG--KNDDDFLNYGISS 229

DB 323 ATRQVNGVYTEEVPTLVLDVDEFTFNSWFCVPGPTNPRLCALAGTAAGHG--KNDDDFLNYGISS 229

QY 379 -----LGRMPTWISPYSKGYIEQYGTDPVT-----GKPAPYSATSVLKT 420

DB 376 SGGIRGPIGLGFRVPCFVLSYSGGLMVHVDREDHTSQLQIGKREGVPPVNLTPRASV 435

QY 421 LGLWLDIEDF--TPRVAHSFSDH-----LIGTLREDAPIALKTPHTFSV 464

DB 436 TGDMTSAFNAAPDPSPFNLDHPVRQLPKVAKCPVNVVGLFNEGLPYRVPYPTTPV 494

## RESULT 4

G70662

probable plcB protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: G70662

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,

Rajandream, M.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: G70662

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-512 &lt;COL&gt;

A:Cross-references: GB:283860; GB:AL123456; NID:g3261681; PIDN:CAB06147.1; PID:g17812

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: plcB



Db 190 AAPSGDGDVHIGNDMDGTIGASGTVDMWRYQ--EGYRSSLWY 247  
 Qy 261 TVYNOTSSNVVPVNF-----FODAVL-----GVLP 287  
 Db 248 LYDAYWYKLEQENYDCNALAFRNFKNAPRDSLDQWRLARGVDQQLKDKVQENTLP 307  
 Qy 288 KFSYINPSCCGTNTNSMHTPGNVSYGEVFKQIYDAIROGPQ-WDKTLLFTTYDETGFGY 346  
 Db 308 QVSWTAVPYCYCE----HPWGPSEGEYIVTRVLEALTSNPEVWARTVFLNLDGDFY 363  
 Qy 347 DIVPPPLAVRPDLNLTVTETAKNGOKYTLHDFR-----LGRMPTWILSPYSKKGXYEQY 401  
 Db 364 DHASAPVPKWDGVLSTVSTAGE---IEVSSGLPGICLGRVPLIAISPWSKGGKV---- 416  
 Qy 402 TDPVTGKAPYSATSUVTKL-----GYLWDIEF-----TPRV 434  
 Db 417 -----SAEVEDHTSVLRFLERRFGWEEENISPRRAVCGDUTSLFDFQAGDQVAPDL 470  
 Qy 435 AHSPSFDHLIGTTLREDA 452  
 Db 471 TNVPSD-----ARKEDA 483

RESULT 7  
 H70662  
 probable plca protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: H70662  
 R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: H70662  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-512 <COL>  
 A:Cross-references: GB:283860; GB:AL23456; NID:g3261681; PID:CAR06148.1; PID:q1781258  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: plca

Query Match 11.4%; Score 286; DB 2; Length 512;  
 Best Local Similarity 28.6%; Pred. No. 2.9e-13;  
 Matches 128; Conservative 40; Mismatches 161; Indels 118; Gaps 20;

Qy 7 LTGLGLASL-GLASPVTSYTSVREAPFGYKPGSKSEIENLKDKVENIYWLILENRSFD 65  
 Db 10 LTGAGAAFLMDWAAPVIEKAYG-----GPCPG-----HLTD-IEHIVLLMQENRSFD 57  
 Qy 66 NILGVRQGLDNPINNPFQNYKNASDPSSGKY---CTQAKDYDSV----- 109  
 Db 58 HFGTLLS-----TNGFNASPAFOQNGWNPMTQALDPAGVTPPRLDTRGP 105  
 Qy 110 -----FNDPDHSTVGNLEFGYTPNNGAIASGVKQVADQSGFLNQLNDYKPLAPEA 163  
 Db 106 FLDGECVNDPEHGWCMHLAW-----NGGA-----ND--NWLPAQA 139  
 Qy 164 TRQ-----VMGYYTEEEVPTLVLDLDETTNSWFSVCPGTPNRLCAL-----AG 210  
 Db 140 TTRAGPYVPLTMGYTRQDPIHLLADTTICDGHCSLLTGTLPNRLYKLSANIDPAG 199  
 Qy 211 TAAGHKNDKDDFLNYGSSKSF-EAANEKGVSWLNVDGNTN-GEFEPDLSFFTYVNOTSR 268  
 Db 200 TDGGPQLVEPGLPQQFSWRIMPENLEDAGVSKVYQKNGLGRFINTPISNNGLVQAFR 259  
 Qy 269 SNWVVEN-----FFQDAVLGLPKFSYINPSCCGTNTNSMHTPGNTGYSYGEV 315  
 Db 260 QAADPRSNLARYGIATPYGDFADVAVRANRLPKVSLVPINT-----LOSEHPALVALGAV 315

Qy 316 FVKQIYDAIROGPQ-WDKTLLFTTYDETGFGYDHYVPPPLA-----VRPNLUYTTETA 366  
 Db 316 SNWTRILILSNPAVWEKIALIVSYDENGGFDDHVTPTTAPPPTGPGFVFVTPNID-AVPG 374  
 Qy 367 KNGOKYTLHDFRLGGRMPTWISPYSK 393  
 Db 375 SGGIRGPL---GLGFRVPCIVISPSYR 398

RESULT 8  
 A36143  
 phospholipase C (EC 3.1.4.3), nonhemolytic - Pseudomonas aeruginosa  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 30-Nov-1990 #sequence\_revision 30-Nov-1990 #text\_change 08-Oct-1999  
 C:Accession: A36143  
 R:Ostroff, R.M.; Vasil, A.I.; Vasil, M.L.  
 J. Bacteriol. 172, 5915-5923, 1990  
 A:Title: Molecular comparison of a nonhemolytic and a hemolytic phospholipase C from  
 A:Reference number: A36143; MUID:91008968  
 A:Accession: A36143  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-692 <OST>  
 A:Cross-references: EMBL:M59304; NID:g151495; PIDN:AAA25968.1; PID:g151496  
 C:Keywords: phosphoric diester hydrolase

Query Match 10.7%; Score 267; DB 2; Length 692;  
 Best Local Similarity 25.2%; Pred. No. 1.1e-11;  
 Matches 114; Conservative 64; Mismatches 181; Indels 94; Gaps 20;

Qy 11 GLLASLGLASPVTSYTSVREAPFGYKPGSKSEIENLKDKVENIYWLILENRSFONILG- 69  
 Db 15 GTVAATVATSMSPSSQAALPAHRRHG-----NLKD-VEHVIVLMQENRSFDHYFGT 67  
 Qy 70 --GVRRG--LDNPNNGPFCNYKNASDPSSGKY-----CTQAKDYDSVNDPDHSTVGN 120  
 Db 68 LKGVRGFGDRNALPLPDQGRVHWKSGKGEILPYHFDTSSTSAQRVDG----- 115  
 Qy 121 NLEFYGTYPNNGAIASGVKQVADQSGFLNQLNDYKPLAPEATQVMGYTTEEVPTLV 180  
 Db 116 -----TPHTWPA-----QQAWNEGRMD---KWLPAK-TERSLGYKEQDIAFOF 156  
 Qy 181 DLVDEFTTENSFSCVPGTPNRLCALGT---AAGHG---KNDDDF---LNYGSSK 230  
 Db 157 ANANFTICDAYHCSFGGTNPRLFLWTGTNDPLQGHGVPVTTNDHDSNGPVEGYTWT 216  
 Qy 231 STFPAANEKGVSWLNVDGNTNGEFPDLS--FFTYVNOTSRSNV-----PVENFFQDA 281  
 Db 217 TYPERLQAAGITWRVYQDMADNFDSDNPLIGFRQYRAAAPDSPLIIVNGLSTWKLDALKRDV 276  
 Qy 282 YLGLPKFSYINPSCCGTNTNSMHP-TGNVSYGEVFKQIYDAIROGPQ-WDKTLLFTTY 339  
 Db 277 LANSLPQVSWL---VAPAKYSEHPGSSPIWCAEYTSWVLDALTANPEVWKSATLLVWF 332  
 Qy 340 DETGFDYDHYVPPPLA--VRPNLUYTTET-----AKNGOKYTLHDFRLGGRMPTWI 388  
 Db 333 DENDGFDHVAAPPAAPSLNKLKDTLRKTTADATLEWHTKGDIRYRNQPYGLGARVPMYVI 392  
 Qy 389 SPYSKKGVIQYGTDPVTGKAPYSATSVLKL 421  
 Db 393 SEWSKGGVNS-----QVFDHTSVIRFL 415

RESULT 9  
 E83230  
 non-hemolytic phospholipase C precursor PA3319 [imported] - Pseudomonas aeruginosa (s  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Sep-2000  
 C:Accession: E83230  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.; Kas, A.; Larbig, K.; L



## RESULT 14

probable phospholipase C - *Streptomyces coelicolor* (fragment)  
C:Species: *Streptomyces coelicolor*  
C:date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
R:Accession: T35551  
C:Seeger, K.J.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajan  
submitted to the EMBL Data Library, July 1999  
A:Reference number: Z21582

Submitted to the EMBL Data Library, July 1999  
A; Reference number: Z21582

A;Accession: T35551

A: Status: preliminary; translated from GB/EMBL/DDP3  
A: Molecule type: DNA

A; Residues: 1-221 <SEE>

A: Experimental source: strain A3(2)  
A; Cross-references: EMBL:AL096825; PIDN:CAB48887.1; GSPDB:GN00070; SCOEDB:5

C;Genetics:

A; Gene: SC0EDB:SC6G3.07

Query Match 5.1%; Score 128; DB 2; Length 221;  
Best Local Similarity 26.4%; Pred NC 0.031.

Matches 60; Conservative 28; Mismatches 83; Indels 56; Gaps

Qv 33 BEGYKPCSKESTENIKKKVENTVMI TI ENPSEDNTI C-- -CVRBOCI DNPTNNCPDECNVV 99

1.  $\frac{1}{2}$  2.  $\frac{1}{2}$  3.  $\frac{1}{2}$  4.  $\frac{1}{2}$  5.  $\frac{1}{2}$  6.  $\frac{1}{2}$  7.  $\frac{1}{2}$  8.  $\frac{1}{2}$  9.  $\frac{1}{2}$  10.  $\frac{1}{2}$  11.  $\frac{1}{2}$  12.  $\frac{1}{2}$  13.  $\frac{1}{2}$  14.  $\frac{1}{2}$  15.  $\frac{1}{2}$  16.  $\frac{1}{2}$  17.  $\frac{1}{2}$  18.  $\frac{1}{2}$  19.  $\frac{1}{2}$  20.  $\frac{1}{2}$  21.  $\frac{1}{2}$  22.  $\frac{1}{2}$  23.  $\frac{1}{2}$  24.  $\frac{1}{2}$  25.  $\frac{1}{2}$  26.  $\frac{1}{2}$  27.  $\frac{1}{2}$  28.  $\frac{1}{2}$  29.  $\frac{1}{2}$  30.  $\frac{1}{2}$  31.  $\frac{1}{2}$  32.  $\frac{1}{2}$  33.  $\frac{1}{2}$  34.  $\frac{1}{2}$  35.  $\frac{1}{2}$  36.  $\frac{1}{2}$  37.  $\frac{1}{2}$  38.  $\frac{1}{2}$  39.  $\frac{1}{2}$  40.  $\frac{1}{2}$  41.  $\frac{1}{2}$  42.  $\frac{1}{2}$  43.  $\frac{1}{2}$  44.  $\frac{1}{2}$  45.  $\frac{1}{2}$  46.  $\frac{1}{2}$  47.  $\frac{1}{2}$  48.  $\frac{1}{2}$  49.  $\frac{1}{2}$  50.  $\frac{1}{2}$  51.  $\frac{1}{2}$  52.  $\frac{1}{2}$  53.  $\frac{1}{2}$  54.  $\frac{1}{2}$  55.  $\frac{1}{2}$  56.  $\frac{1}{2}$  57.  $\frac{1}{2}$  58.  $\frac{1}{2}$  59.  $\frac{1}{2}$  60.  $\frac{1}{2}$  61.  $\frac{1}{2}$  62.  $\frac{1}{2}$  63.  $\frac{1}{2}$  64.  $\frac{1}{2}$  65.  $\frac{1}{2}$  66.  $\frac{1}{2}$  67.  $\frac{1}{2}$  68.  $\frac{1}{2}$  69.  $\frac{1}{2}$  70.  $\frac{1}{2}$  71.  $\frac{1}{2}$  72.  $\frac{1}{2}$  73.  $\frac{1}{2}$  74.  $\frac{1}{2}$  75.  $\frac{1}{2}$  76.  $\frac{1}{2}$  77.  $\frac{1}{2}$  78.  $\frac{1}{2}$  79.  $\frac{1}{2}$  80.  $\frac{1}{2}$  81.  $\frac{1}{2}$  82.  $\frac{1}{2}$  83.  $\frac{1}{2}$  84.  $\frac{1}{2}$  85.  $\frac{1}{2}$  86.  $\frac{1}{2}$  87.  $\frac{1}{2}$  88.  $\frac{1}{2}$  89.  $\frac{1}{2}$  90.  $\frac{1}{2}$  91.  $\frac{1}{2}$  92.  $\frac{1}{2}$  93.  $\frac{1}{2}$  94.  $\frac{1}{2}$  95.  $\frac{1}{2}$  96.  $\frac{1}{2}$  97.  $\frac{1}{2}$  98.  $\frac{1}{2}$  99.  $\frac{1}{2}$  100.  $\frac{1}{2}$

Db 33 PANHRSGSIED-----VEHIWLMQENRSFDHYFGTLRGVRGFGDHPV----- 76

Qy 90 NASDPSSGKYCTQAKDYDSVFNDPDHVSVTGNNLEFYGTYPNNGAIASGVVADQSGFLN 149

77  
77

EST COMMUNITARIUM IN ITALIA E IN EUROPA

QY 150 AQ--LND--YPKLAPEEATRQVMGYYTEEEVPTLVLDLVEFTTFNSWFSCVPGPTNPRL 205

Db 120 GODAYHDAKYDRWLPAGKT-TT MAYLTREDIPEHYALADTFTVCDAYHCSEFIGSTDPNRY 178

QY 208 CALAGTAAAGHGKND-----DDFLNYGTSRSKSFEEAANERGVSW 243

Db 179 YLW----SGHTGNDGAGGGPVLGNDELGYDWTYP--ERLEAAGVSW 219

RESULT 15

acid phosphates / EC 2 1 2 2 1 - 2000/11/06 640000

C;Species: *Aspergillus ficuum*

```
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 12-Sep-1997
C;sequence_revision: 702380. 502242
```

R; Ehrlich, K.C.; Montalbano, B.G.; Mullaney, E.J.; Dischinger Jr., H.C.; Ul

Biochem. Biophys. Res. Commun. 204, 63-68, 1994

A; Reference number: JC2389; MUID:95032134

A;Accession: JC2389

A:Residues: 1-417 <EHR1>

A;Cross-references: GB:L20566; NID:g304094; PID:g304095

A;Accession: PC2243  
A:Molecule type: protein

A; Residues: 32-180; 204-220; 252-297; 325-413 <EHR2>

C;Genetics:  
A:Gene: AFPba

A; Introns: 64/3

1



C:Keywords: glycoprotein; phosphoric monoester hydrolase

F:215-216/Region: catalytic; #status predicted

F:122,187,209,218,333,383/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 4.8%; Score 119.5; DB 2; Length 417;  
Best Local Similarity 20.6%; Pred. NO. 0.32;  
Matches 74; Conservative 43; Mismatches 137; Indels 105; Gaps 15;

```
QY 155 YPKLAPEEATROVMGYT-----EEVPPLVD---LV 183
||| ||| : : : ||| : : : ||| : : : |||
Db 26 YPSLAELRAAQATVQPSVSNVKGITFNRFNWIWLENTDFDAAAATEHLPVLAKKGLLL 85
||| ||| : : : ||| : : : ||| : : : |||
QY 184 DEFTTFNSMFCVPGPTNPRLCALAGTAAGHGKNDLDFLNYGISSKSLFEAAAEKGVSW 243
:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 86 NNF-----W--AVTHPSEP-NYCRHPLPLGDTFGMDNDDFHQIPSNVSTIADLFDTKNIAW 137
||| ||| : : : ||| : : : ||| : : : |||
QY 244 -----LNYDGTNGEFEPDS-----LFFTYVNTSR-SNVVPVENFQDAYL 283
||| ||| : : : ||| : : : ||| : : : |||
Db 138 GEYQEGLPYGGYRYPESGANDYVRNRNPLILFDSVTEDALRLRQIKNFSSFYDDLEN 197
||| ||| : : : ||| : : : ||| : : : |||
QY 284 GVLKFSYINPSCGNTNSMHPITGNVSYGEVFVKQIYDAIRQGPQMDK-TLLFITYDET 342
||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 198 HRLPQYMFITPNN-----TNDGHD-TNITSGDWTGFLSELLENDFYTKDTLIMLDET 252
||| ||| : : : ||| : : : ||| : : : |||
QY 343 GGFY-----DHVPPPLAVRPDLTYT-----ETAKNCQKYLHFDRLGGRMPT 385
||| ||| : : : ||| : : : ||| : : : |||
Db 253 GTYEIGNNIYTFLLGGAVPDDLGLTKDDTFYTHYSVIASLSTNWGLPSLGRWDCGANLFS 312
||| ||| : : : ||| : : : ||| : : : |||
QY 386 WVI-----SPYSKKGYIEQYGTDPVTGKPAYSA-TSVLRTL 421
||| ||| : : : ||| : : : ||| : : : |||
Db 313 WLAKKTCYVNYEVDTSNLYMNETHWGPLSDDDYSEYAGWVPVPTDASCAGNGILSTV 371
||| ||| : : : ||| : : : ||| : : : |||
```

Search completed: February 26, 2001, 16:21:42  
Job time: 41382 sec





374 ACAACGGCTTGCAGATGAA..... : 399

173 GluGlucIValProThrLeuAlaAspLeuValaspGluPheThrThrPh 189  
          |||||     |||::       :  
394 .....TCATCGCATGGTTGCCAAGAAGGTATTCTTT 428  
  
189 eAnSerTipHeSerCysValProGlyProThrAsnProAsnArgLeu 206  
          |||||     |||::       :  
429 GACTTAATACTGGTCT...TTGACTCAACCCTTGACAACAACTACTTGG 475  
  
206 ysAlaLeuAlaGlyThrAlaLaGlyHisGlyLysAsnAspAspPhe 222  
          :|||:: |||:: |  
476 CCTGTGCTGGTAGCTAC....TTTCCTGTGAATGATGACAGGTTCC 519  
  
223 LeuAsnTyrlYlleSerSerLyserlSerrlePheGlnAlaAlasnGly 239  
          ::::: ::::|::|::|::|::|  
520 ATCTCCAATGCCATCGAATGTCTCCAACATCGTTGATTGTTAGATACCAA 569  
  
239 sGlyValSerTrp.....LeuAsnTyrAspGlyThr 250  
          :|||:: |||:: |  
570 GAACATTTCTTGGGCAGAATATCAAGAACATATTCATCACACTGGTTCCC 619  
  
250 snGly.....: 251  
          ::::|  
620 AGGCTTCAACTACTCTACACGAGAAACCTTCGCCAACGACTATGTGAGA 669  
  
252 GluPheGluProAspSerLeuPhePheThryrValasnGlnThrSerAr 268  
          :|||:: |||:: |  
670 AHCACACACCCATTGATTATTTCGATACAGTTATTTCGACAAAGTCACG 719  
  
268 g...SerAsnValValProVaIGluAsnPhePheGlnAspAlaTyrLeug 284  
          |||||     |||::       :  
720 TTTGGCTAACATTAAGAGTTTTGGAGCACTTCAACACTGACTTCAAGAACA 769  
  
284 lyValLeuProLysPheSerTyrilleAsnProSerCysGlylThrasn 300  
          :|||:: |||:: |  
770 AGACTCTACCACAAATGCTTTCATTACTCCAAACATG..... 807  
  
301 ThrAnsSerMetHispProthrClYasNvalSerTyrglyUValPheVa 317  
          |||||     |||::       :  
808 ACCAATGATGTCATCATCA...AATATCAAGTTGCCGGTCACTGGTTC 854  
  
317 llysGliniletyrAspAlalaIrAgIngInglyProdInTraSpLys...T 333  
          :|||:: |||:: |  
855 TAATCTTTCTTCGAACCATTTGTCAGTGATGATTACTTCATCAAGAGATA 904  
  
333 hrLeuLeuPheIlleThyrrasp.....gluthRgilyGlyphefyrAsp 347  
          :|||:: |||:: |  
905 CTTTAGCTCTTTAACTTTGATGAANAACCAACTATGTTGATCAAGAAC 954  
  
348 HisValIProProLeu....AlavalargProAspnlenLeuThy 362  
          |||      |||     |||::|  
955 AAGGTTTTCTCTATCTATTAGTGGGTGTTATCCACGACCACCTGAAGG 1004  
  
362 rThrGluThrAlalaysanGlyGlnLysTyrThrLeuIIePheAspArgL 379  
          |||      |||     |||::|  
1005 TACT.....CTGACGATACATCTTACGACCATTTACTCTCAAT 1042  
  
379 euGlyelyargmetProthrtp 386  
        |||:: |||:: |  
1043 TGCTACTGTTGAGGCTAAGTGG 1065

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1997.DAT:T44068

seq\_documentation\_block:

ID T44068 standard; cDNA; 4360 BP.

XX  
AC T44068;

DT 27-FEB-1997 (first entry)

Rat von Ebner's gland protein Ebnerin cDNA.  
XX  
XX

```

Ebnerin: von Ebner's gland; saliva; salty taste perception; ds.
Rattus rattus.
Key Location/Qualifiers
CDS 94..3966
FT /tag= a
XX
XX WO9639513-A2.
XX
XX 12-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US08966.
XX
XX 06-JUN-1995; 95US-0470350.
XX
XX (UOJO ) UNIV JOHNS HOPKINS SCHOOL MED.
XX
XX LI X, Snyder SH;
XX
XX WPI; 1997-043135/04.
XX
XX P-PSDB; W07609.
XX
XX Von Ebner's gland derived polypeptide, ebnerin - useful to develop
XX prods. for investigating taste perception
XX
XX Claim 11; Page 22-27; 48pp: English.
XX
XX A cDNA clone (T44068) codes for a novel secreted protein, Ebnerin
XX (W07609), which is formed in the ducts of von Ebner's gland and
XX secreted into fluid bathing the taste buds contained in the taste
XX papillae. It was obtd. by PCR amplification of rat circumvallate
XX papillae cDNA using primers (T44069-70) based on the amiloride
XX sensitive sodium channel, and use of a PCR product to screen a
XX papillae cDNA library. Ebnerin nucleic acids can be used to
XX produce recombinant protein and as probes to detect Ebnerin
XX polynucleotides.
XX
XX Sequence 4360 BP; 1051 A; 1099 C; 1103 G; 1107 T; 0 other;
XX
alignment_scores:
Quality: 122.00 Length: 303
Ratio: 0.878 gaps: 16
Percent Similarity: 45.875 Percent Identity: 23.762
alignment_block:
US-09-426-072-2 x T44068 ..
Align seg 1/1 to: T44068 from: 1 to: 4360
86 CysAsnTyr.....LysAsnAlaSerAspPr 94
|||||
2758 TGCAACTATGACTATATAGAGATTTTGTGGCCCCCACCACAGTTCACCC 2807
94 oSerSerGlyLysTyrCys.....ThrGlnAlaL 104
|||||
2808 TCTCATTTGCCGGGTTTGTGTGGGGCCATGGGCTCTTTCACTTCAACAT 2857
104 ySAspTyrAspSerValPheAsnAspProAspHisSerValThr..... 118
|||||
2858 CCAACTTCATGTCAGTTCGCTTCACCACTGATCAGAGTTTGACAATAATACCAACAT 2907
119 GlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAsnAsnGlyAlaI 135
|||||
2908 GGGTTCCGGGCTGACTACTACTACTAGACTTTTGACAATAATACCAACAT 2957
135 eAlaSerGlyLysValAlaAspGlnSerGlyPheLeuAsnAlaGlnL 152
|||||
2958 CCTTGT.....C 2965
152 euAsnAspTyrProLysLeuAlaProGluAlaThrArgGlnValMet 168
|||||

```





## alignment\_block:

US-09-426-072-2 x Q42754

Align seg 1/1 to: Q42754 from: 1 to: 2062

```

110 PheAsnAspProAspHisSerValThrGlyAsnAsnLeu.....122
|||||
627 TTCAATCCAGCGGACTGGCAGTTAATATCCAAACACATGACAGAAATAAA 676
123 .....GluPheTyr.....125
|||||
677 CTTAGTAGTCTTTTGAACAGAAATATTCATCTAGTACTATAAAACAATTA 726
126 .....GlyThrTyrThrPro.....AsnAsnGlyAlaIle 135
|||||
727 CAGAAATCAGCAACCTCACCACCATCCAAAATATATAATATGATCTAACT 776
136 AlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLe 152
|||||
777 GCAAGCTTAATGGTGGCAGTACAGACCAATAACACACTT.....815
152 uAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnValMetG 169
|||||
816 ....CCATACACACAGCAGCAGCAGTAAAGTGAACA.....CTTG 852
169 lyTyrTyrThrGluGluGluValProThrLeuValAspLeuValAspGlu 185
|||||
853 GTTTTAT.....CCATGGTTACCTACA.....AAA 878
186 PheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsnPr 202
|||||
879 CCAACTCAATACAGATATTACCTATCATGCAGAAACCTTAATCCACC 928
202 oAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsnA 219
929 A.....ACATACACTGGCAATCACAACAA 954
219 spAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePheGluAla 235
|||||
955 TAACAGACTCAATACAAACAGGACTACAGTGCATTTATGTTCTACACA 1004
236 AlaAsnGlyLysGlyValSerTrpLeuAsnTyrAspGly.....Th 249
1005 ATAGAAAATGCGACTACCAATTCATCTCTTAAGAACTGGAGATGAATTC 1054
249 rAsnGly.....GluPheGluProAspSerLeuPhePheThrTyrVala 264
1055 CACAGGAATATATCACTTTGACACAAAACCATTAATAATTAACCTCAT 1104
264 sGlnThrSerArgSerAsnValValProValGluAsnPhePheGln... 279
1105 GCAACAAACAGACTCTCTAGGACTGCTCCCAAAAGTACTAAGTGAACCT 1154
280 .....AspAlaTyrLeuGlyValLeuProLysPheSerTyrIl 292
1155 ACCACAGAGAGAGACCAACCCAGGAACACTACCA.....1190
292 eAsnProSerCysCysGlyThrAsnThr...AsnSerMethHisProThrG 308
1191 .....GGAGCTACACAGAAAGGTTATCACCACAA 1224
308 lAsnValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIle 324
1225 TTAATAATAGCTACACAGAAGCA.....ACAGCACTT 1256
325 ArgGlnGlyPro...GlnTrpAspLysThrLeuLeuPheIleThrTyrAs 340
1257 AGGCCAGCTCAGTAGGATATATACACCATACATGAATTTTGACTACTC 1306
340 pGluThrGlyGlyTyrAspHisValProProLeuAlaValArgp 357
1307 CAATGGTGGACCACTTTCTA.....ACTCCTATAGTACCAACAG 1344

```

```

357 roAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThr 373
|||||
1345 CAGACACACAATATTATGATGATGATGAACCAATGGTGTATAAAGATTTACA 1394
374 Leu.....HisPheAspArgLeuGlyGlyArgMetProTh 385
|||||
1395 ATGGTTTACCACATGGAGACTTAACACACATCTTCACAAGAGCTAGANAAG 1444
385 rTrpValIleSerProTyrSerLysLysGlyTyrIle.....GluGlnT 400
|||||
1445 ATACATATTCAATCCACAAAGTAATGTGGAAGAGCTCCAAAGCAACAAT 1494
400 yGly.....401
1495 TTAATCAACAGGCACCACTAAACCTAGAAAATACAAATATATGAACACTT 1544
402 .....ThrAspProValThrGlyLysPro.....409
1545 TTACCTTTCAAGTCCCAATAGGAGGAATCTAACAAGCATTTTCATGAATAC 1594
410 .....AlaProTyrSerAla.....ThrSerValLeuL 419
1595 ACTCAATACATATGGACCATTAACAGCACTAAACAATACTGCACCTGTAT 1644
419 ysThrLeuGlyTyrLeuTrpAspIleGlu.....AspPheThrPro 432
1645 TTCCAATGTCAATATGGGATAAAGAACTTGATACAGACTTAAACACCT 1694
433 ArgVal 434
1695 AGACTA 1700

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/NA1999.DAT:X60600

seq\_documentation\_block:

ID X60600 standard; DNA; 3126 BP.

XX X60600;

XX 27-JUL-1999 (first entry)

XX S. cerevisiae acidic trehalase (ATH1) gene sequence.

XX Acidic trehalase: ATH1 gene; gene manipulation; baker's yeast; dough; fermentation; baking; freezing; bread; trehalose degradation; ss.

XX Saccharomyces cerevisiae.

XX EP921190-A2.

XX 09-JUN-1999.

XX 04-DEC-1998; 98EP-0403046.

XX 08-DEC-1997; 97JP-0352016.

XX (NORQ ) NAT FOOD RES INST MIN AGRIC.

XX (ORIY ) ORIENTAL YEAST CO LTD.

XX Iyo C, Mori K, Nakajima R, Shima J, Suzuki Y, Takano H;

XX Watanabe H;

XX WPI; 1999-315405/27.

XX P-PSDB; Y16788.

XX New acid trehalase (ATH1) gene-disrupted haploid yeast, useful for preparing high-sugar dough resistant to freezing

XX Examples; Page 16-23; 57pp; English.

XX The invention relates to an acid trehalase (ATH1) gene-disrupted haploid yeast produced by gene manipulation, where the diploid is a practical

CC

baker's yeast. The new yeast strain is useful in dough which, upon subsequent fermentation and baking after a period of freezing, produces a high quality bread or sweetened bun. The new yeasts have a reduced ability of trehalose degradation during fermentation and thus an increased intracellular osmotic pressure, providing resistance to long term freeze-storage when used in low and high-sugar dough. The dough makes better bread than dough using prior art baker's yeast. The new strain and A1H1 non-disrupted gene strains were used to prepare dough, and the new strain showed resistance to freezing. The present sequence represents an A1H1 gene sequence.

```

alignment_scores:
  Quality: 109.50      Length: 452
  Ratio: 0.956        Gaps: 24
  Percent Similarity: 43.584  Percent Identity: 19.469

alignment_block:
  US-09-426-072-2 x X60600
  ..
  Align seg 1/1 to: X60600 from: 1 to: 3126

      22 ValThrSerGluTyr.....ThrSerValArgGluAlaProPheGl 35
      ||| ||||| ||||| :|||:|||| ||| |||:
    1336 GTCCGATTACGAAATACCATATTACGCTTGATGCGTATGGCCTCTTTC 1385
      35 yTyLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValG 52
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1386 CATATACTTGAATGGACACGAAGGATTGATGAC.....GAGTATCTGA 1429
      52 luAsnIleValTrpLeuIleLeuGluAsnArgSer..... 63
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1430 GATATCTACATGCCCAATATCAAAACGACGCCCAATTTTTTACTGCT 1479
      64 .....PheAspAsnIleLeuGlyGlyValArgGlnGlyLeuAs 77
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1480 TATGTTAAGTACAATCTTCCCTAGGATTGTATCAACATATAATTTCAC 1529
      77 pAsnPro.....IleAsnAsnGlyPProPheCysAsnT 88
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1530 AGATCCCGACGAGTTTGCTAATCATCATCAATACGGGGCTTTC..... 1572
      88 yLysAsnAlaSerAspProSerSergLysTyrcysThrGlnAlaLys 104
      ||||| |||: |||: |||: |||: |||: |||: |||: |||:
    1573 ...ACGAATGCTGGCATTAAACACTCTTAAGATGGCGCAACAGACATTGGC 1620
      105 AspTyraSPSerValPheAsnAspProAspHisSerValThrGlyAsnAs 121
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1621 AATCATCTCGCGGNGTCGTGACCCCAATGAGTGAATTTCCAAAGA 1670
      121 nLeuGluPheTyrgLyThrTyThrProAsnAsnGlyAlaIleAlaSerG 138
      ||| :||| |||: |||: |||: |||: |||: |||: |||:
    1671 TAIT.....TATATCCCTAGA.....TCCTCAT 1693
      138 lYlYsValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAsp 154
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1694 CTAACATCACTTTTGGAAATATCTGCT.....ATGAATAGC 1728
      155 TyTProLysLeuAlaProGluGluAlaThrArgGlnVal.....MetGl 169
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1729 TCAGTGGAGATTAAACGCGGATGTGACTTTAATGGTTTACCCACTTGG 1778
      169 yTyTyThrGluGluValProThrLeuValaspLeuValAspGluP 186
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1779 ATATATCAATGATCAATCCAVA.....TTGAACAATGCAATTAAGATC 1822
      186 heThrThrPheAsnSerTrpPheSerCysVal...ProGlyProThrAsn 201
      :|||: |||: |||: |||: |||: |||: |||: |||:
    1823 TTTATTATTATCAGAAAGACAGTCTCGGCTCGGCTCGCAATGACATAT 1872
      202 ProAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGly..... 216
  
```

1873 CGGGTTTTTGTGGCCGAGCTGCTGGT...CTGCTGAATCAGGCTCTTC 1919  
216 .....  
1920 TTCTCAAGATTACTTATATAAATCGGTCCTTCATCATCTACGGGCTCTTC 1969  
217 .....LysAsnAspAspPheLeuAsnTyrGlyIle 227  
1970 TCGCTCAATTTAGTGAGCAATCAGACGACAACCTTTTAAACAAACGGATT 2019  
228 SerSerLysSerIle.....PheGluAlaAl 236  
2020 ACCCAGCCAGCATTTCCCTTTTAAACAGCTAACGGTGATTCTTCACAG 2069  
236 aAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyLup 253  
2070 CAATCTGTGTGGTTAACAGGAATCCGATAT.....TCATTATGAG 2110  
253 heGluProAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSer 269  
2111 TTGATCCGAGTACT.....AAAAAATAACCGTTGTGAAGGTC 2151  
270 AsnValProValGluAsnPhePheGlnAspAlaTyrLeuGlyValIle 286  
2152 AAT.....CCAATAGAA.....CTACCGTTGCT 2174  
286 uPro.....LysPheSerTyrIleAsnProSerCysC 297  
2175 CCCTGGTGATCGCTATTAGAACTTCAAAATATATGAC..... 2214  
297 ysGlyThrAsnThrAsnSerMethHisProThrGlyAsnValSerTyrGly 313  
2214 .....  
314 GluValPheValLysGlnIleTyrAspAlaIleArgGlnGlyProGlnTr 330  
2215 .....CAAGTTTATAGAT..... 2226  
330 pAspLysThrLeuPheIleThrTyrAspGluThrGlyGlyPheTyrA 347  
2227 .....ATAAATATGACGACCACAATGGTACCATG 2257  
347 sPhis.....ValProProProLeu 353  
2258 TTCATAAATCAGGAGATGTTCTTCCATATCAAGATACCAACAGATCT 2307  
354 AlaValArgProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGl 370  
2308 CTAATACATGACGAGGATATCAACTTCTATAATGGTTCGCAAAACGAAG 2357  
370 nLysTyrThrLeuHis.....PheAspArgLeuGlyGlyArgMet 384  
2358 AAAACCAAACTAGACCGTAGAGACGTCGACCGCTGTTGGTGATCCAATG. 2406  
384 roThrTrpValIleSerProTyrSerLysGlyTyrIleGluGlnTyr 400  
2407 .....AGGATGGATAGGTAT 2421  
401 GlyThr 402  
2422 GGTACC 2427

seq\_name: /cgn2\_2/gcgcdata/geneseq/geneseq/NA1999.DAT.X60602

seq\_documentation\_block:

ID X60602 standard; DNA; 4292 BP.

XX  
AC X60602:XX  
DT 27-JUL-1999 (first entry)

DE *S. cerevisiae* ATH1 gene coding region as cleaved with URA3 gene.





```

||||
3424 TTCAATAATCAGGAGATCTTCCTATTATATAAAGATACCAAAACAGATCT 3473
|||||
354 AlaValArgProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyL 370
|||||
3474 CTATACATGACCGAGGATCACTTCTATATAGTTCGCAAAACGAAAG 3523
|||||
370 nLysTyrThrLeuHis.....PheAspArgLeuGlyGlyArgMetP 384
|||||
3524 AAAACCAATCTAGAGCGTAGAGACGCGGTGTGTGTGATCCAATG. 3572
|||||
384 rofhrtrpValIleSerProTyrSerLysGlyTyrIleGluGlnTyr 400
3573 .....AGCATGGATAGTAT 3587
401 GlyThr 402
|||||
3588 GGTACC 3593

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1998.DAT.X14547

seq\_documentation\_block:

ID X14547 standard; DNA; 8748 BP.

XX AC X14547;

XX DT 31-MAR-1999 (first entry)

XX DE H. pylori GHPO 1484 gene.

XX KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
KW peptic ulcer disease; ss.

XX OS Helicobacter pylori.

XX FH Key Location/Qualifiers

XX FT CDS 16..8697

XX FT FT /\*tag= a

XX PN W09843478-A1.

XX PD 08-OCT-1998.

XX PF 01-APR-1998; 98MO-US06371.

XX PR 29-JUL-1997; 97US-0902615.

XX PR 01-APR-1997; 97US-0833457.

XX PR 24-JUN-1997; 97US-0881227.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX FA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX XX WPI; 1998-542293/46.

XX DR P-PSDB; W98828.

XX XX New isolated Helicobacter polynucleotides - used to develop products  
XX FT for the diagnosis, prevention and treatment of Helicobacter  
XX PT infections and gastrointestinal diseases

XX XX Claim 1: Page 1827-1840; 2054pp; English.

XX CC This sequence represents a polynucleotide of the invention. It was  
XX CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.  
XX CC the polypeptides can be used for preventing or treating Helicobacter  
XX CC infections, and gastroduodenal diseases associated with these  
XX CC infections, including acute, chronic, and atrophic gastritis, and peptic  
XX CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used  
XX CC for the production of antibodies. The products can also be used for  
XX CC detection and diagnosis.

XX SQ Sequence 8748 BP; 2670 A; 1780 C; 1836 G; 2462 T; 0 other;

```

alignment_scores:
  Quality: 108.00      Length: 455
  Ratio: 0.527        Gaps: 28
  Percent Similarity: 45.055  Percent Identity: 22.198

alignment_block:
US-09-426-072-2 x X14547 ..
Align seg 1/1 to: X14547 from: 1 to: 8748

1 MetLysSerThrAlaLeuLeu.....ThrGlyLeuGlyLeuLeuAl 14
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
836 TTGAAGAGACGCCACCTTTTACGGGGGGAATGGAATGGGGGACTTAT 885

14 aSerLeu.....GlyLeuAlaSerProValThrSerGluTyrThrSerV 29
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
886 ACCTTTAATAAAGAGTTTAGCGCTACCAATAACACCGCCTTTTAGTAGCGG 935

29 aArgGluAlaProPheGlyTyrLysProGlySerLys.....Glu 42
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
936 TAGT.....TTTAATTTTAAAGGTGAAGCTCTTTTAAATGGTACT 975

43 SerIleGluAsnLeuLysAspLysValGluAsnIleValTrpLeuIle 59
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
976 TCGTTTAGTAACGCTCTTATATACCTTTTGACAAT.....CAAGCCACTTT 1019

59 uGluAsnArgSerPheAspAsnIleLeuGlyGlyValArgGlnGlyL 76
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1020 CCAAAACAGCTCCCTTTAAAGGG.....GGGACTTTTACTTTTAAATACC 1063

76 euAspAsnProIleAsnAsnGlyProPheCysAsnTyrLysAsnAlaSer 92
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1064 AACTAATCCCACTAACACGCTCAGCACCCCAATTCAAAACAGCTCT 1113

93 AspProSerSerGlyLysTyrCysThrGlnAlaLysAspTyrAspSerV 109
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1114 TTTAGTGGTAACGCTACCACTCTTAAGGGCTTTGTGAATTTCCAGCAAGC 1163

109 lPheAsnAspProAspHisSerValThrGlyAsnAsnLeuGluPheTyrG 126
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1164 CTTTAAACAAATCAAAACACCACTAACGATCCAAACAGCTCTCTTT...A 1210

126 lyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValValAla 142
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1211 ATAACGCCACTTTTAAACAAT.....ACCGTAAATCACTATA 1248

143 AspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeuAl 159
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1249 GAAAA.....GATCGAGTTTAAATAC..... 1272

159 aProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGluGluVal 176
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1272 ..... 1272

176 aProThrLeuValAspLeuValAspGluPheThrThrPheAsnSerTrp 192
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1273 .....ACGACATTCACACT... 1287

193 PheSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeuAl 209
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1288 .....TCTGTGTATACAAACAACATGAGTGTATACCGG 1319

209 aGlyThrAlaAlaGlyHisGlyLysAsnAsp.....AspAspPheL 223
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1320 TGGCGTACTTTAAGC...GGTAAATAGCTTGAAAAATGGCTCAACCC 1366

223 euAsnTyrGlyIleSerSerLys..... 230
  :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1367 TTGATTTTGGG...AGTTCTAAATCACTCTCGCTCAAGGGACGACTTTC 1413

231 SerIlePheGluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAs 247

```

```

1414 AACCTCACAGTTAGGACGACGAGCGTAACGATTTTAAATCTAG 1463
      :::::  ::  :::::|||||:::|||||  ::  :::::  ::
247 pGlyThrAsnGlyGluPheGluProAspSerLeuPhePheThrTyrVala 264
      :|||  :::::  ::  :::::|||||  ::  :::::  ::
1464 CGGTGGG.....ATCATTATATAGTA 1483
      :::::  ::  :::::|||||  ::  :::::  ::
264 sn.....GlnThrSer 267
      ||  :::::  ::  :::::|||||  ::  :::::  ::
1484 ACCTTTAAACCATGCAATCAACGCGTTGACAGTCCCTTAAACACGAC 1533
      :::::  ::  :::::|||||  ::  :::::  ::
268 ArgSerAsnValProValIcuAsnPheGlnAsp..... 280
      :::::  ::  :::::|||||  ::  :::::  ::
1534 GAAAGCCTTTCAAAATCCG...CAAAGTTTCCTCAAGGTTCGCGGATAT 1580
      :::::  ::  :::::|||||  ::  :::::  ::
281 ....AlaTyrLeuGlyValLeuProLysPheSerTyrIleAsn..... 293
      :::::  ::  :::::|||||  ::  :::::  ::
1581 AATCATTACAAATGGGTACCGGCGACGTTTGAATGAAACGCTCAA 1630
      :::::  ::  :::::|||||  ::  :::::  ::
294 .....ProSerCysCysGlyThrAsnThrAsn 302
      :::::  ::  :::::|||||  ::  :::::  ::
1631 CATCTAAACCCACTGACTCTCGCCCTCTAAATCCTCTACAACTCTACG 1680
      :::::  ::  :::::|||||  ::  :::::  ::
303 SerMetHisProThrGlyAsnValSerTyrGlyGluValPheValLysG 319
      :::::  ::  :::::|||||  ::  :::::  ::
1681 CAAGTCTATCAAGTGGT...TACAAATAGGGGATCTATCTACAACT 1727
      :::::  ::  :::::|||||  ::  :::::  ::
319 n.....IleTyrAspAlaIleArgGlnG 327
      |||  :::::  ::  :::::|||||  ::  :::::  ::
1728 GCAAGAACTTTTCAGCCAAATTCATTTATTCAGGCTTTAGAGAGCG 1777
      :::::  ::  :::::|||||  ::  :::::  ::
327 lyProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGly 343
      |||  :::::  ::  :::::|||||  ::  :::::  ::
1778 GG.....ACTAC..... 1785
      :::::  ::  :::::|||||  ::  :::::  ::
344 GlyPheTyrAspHisValProProProLeuAlaValArgProAspAsnLe 360
      :::::  ::  :::::|||||  ::  :::::  ::
1786 .....ACGCCACCCCTGTCAATT..... 1803
      :::::  ::  :::::|||||  ::  :::::  ::
360 uThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheA 377
      :::::  ::  :::::|||||  ::  :::::  ::
1804 .....AACGGCTCCAAATTTGACTTATCCGCTT 1831
      :::::  ::  :::::|||||  ::  :::::  ::
377 spArg...LeuGlyGlyArgMetProThrTrpValIleSerProTyrSer 392
      :::::  ::  :::::|||||  ::  :::::  ::
1832 CAAATTATATCAATGCTGACATGCCT...TGG.....TATGAC 1866
      :::::  ::  :::::|||||  ::  :::::  ::
393 LysLysGlyTyrIle 397
      |||  :::::  ::  :::::|||||  ::  :::::  ::
1867 CATAAATATTACATC 1881
      :::::  ::  :::::|||||  ::  :::::  ::
seq_name: /cgn2_2/cgdata/geneseq/geneseq/NA1998.DAT.V52091
seq_documentation_block:
ID V52091 standard; DNA; 8748 BP.
XX AC V52091;
XX AC
XX 09-NOV-1998 (first entry)
XX AC
DE Helicobacter polypeptide GHPO 1484 DNA.
XX GHPO 1484; infection; therapy; diagnosis; vaccine; gastritis;
XX ulcer; ss.
XX Helicobacter pylori.
XX Key Location/Qualifiers
XX CDS 16..8697
XX /*tag= a
XX W09821225-A1.
XX

```

```

PD 22-MAY-1998..
XX
PF 14-NOV-1997; 97WO-US21353.
XX
PR 29-JUL-1997; 97US-0902615.
PR 14-NOV-1996; 96US-0749051.
PR 01-APR-1997; 97US-0831309.
PR 01-APR-1997; 97US-0833457.
PR 01-APR-1997; 97US-0834705.
PR 24-JUN-1997; 97US-0881227.
XX
(HUMA-) HUMAN GENOME SCI INC.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C;
PI Odenbreit S, Tomb J;
XX WPI; 1998-297855/26.
DR P-PSDB; W71556.
XX
Helicobacter polynucleotide and polypeptide sequences - useful to
treat or prevent gastrointestinal infection
XX
Disclosure; Page 317-330; 362pp; English.
XX
This polynucleotide encodes Helicobacter pylori polypeptide
GHPO 1484 (see W71556). 85 Polynucleotides (see V52009-V52093)
encoding claimed Helicobacter polypeptides (see W71474-W71558) have
been identified in the H. pylori genome using a transposon shuttle
mutagenesis method in which TnMax9 mini-blam was used for
insertional mutagenesis of an H. pylori gene library established in
Escherichia coli. The invention provides: methods for producing
these polypeptides in recombinant host systems, and related
expression cassettes, vectors and transformed or transfected host
cells; live vaccine vectors that contain polynucleotides of the
invention; therapeutic and/or prophylactic methods involving
administration of polynucleotide molecules (either in naked form
or delivered with a delivery vehicle), polypeptides or monospecific
antibodies; methods for detecting the presence of Helicobacter
in samples using e.g. polynucleotide molecules; and methods for
purifying polypeptides of the invention.
XX
SQ Sequence 8748 BP; 2670 A; 1782 C; 1834 G; 2462 T; 0 other;

alignment_scores:
Quality: 109.00 Length: 455
Ratio: 0.527 Gaps: 28
Percent Similarity: 45.055 Percent Identity: 22.198

alignment_block:
US-09-426-072-2 x V52091 ..
Align seg 1/1 to: V52091 from: 1 to: 8748

1 MetLysSerThrAlaLeu.....ThrGlyLeuGlyLeuLeuAl 14
:::|||||  ::  :::::|||||  ::  :::::  ::
836 TTGAAGAGAACGCCACTTTTAGCGGGCGAAATGGAATGGGGGACTTAT 885
14 sLeu.....GlyLeuAlaSerProValThrSerGluTyrThrSerV 29
|||  :::::  ::  :::::|||||  ::  :::::  ::
886 ACCTTTAATAAGAGTTTAGCGCTACCAATAACACCGCTTTAGTAGCGG 935
29 aLArgGluAlaProPheGlyTyrLysProGlySerLys.....Glu 42
||  :::::  ::  :::::|||||  ::  :::::  ::
936 TAGT.....TTTAAATTTAAAGGTGTAAGCTCTTTTAAATGGTACT 975
43 SerIleGluAsnLeuLysAspLysValIcuAsnIleValTrpLeuIle 59
|||:|||||  ::  :::::|||||  ::  :::::  ::
976 TCGTTTAGTAACGCTTCTTATACTTTTGACAAT.....CAAGCCACTTT 1019

```

```

59 uGluAsnArgSerPheAspAsnIleLeuGlyGlyValArgArgGlnGlyL 76
   :||||| :||||| :|||
100 CCAAAACAGCTCCTTTAAATGGG.....GGGACTTTTACTTTTAAATAACC 1063
   :||||| :||||| :|||
76 euAspAsnProIleAsnGlyProPheCysAsnTyrIleAsnAlaSer 92
   :||||| :||||| :|||
1064 AAATANTCCAACTAACACGCTCAGCACCCCAATTCACAAACAGCTCT 1113
   :||||| :||||| :|||
93 AspProSerSerGlyLysTyrCysThrGlnAlaLysAspTyrAspSerVa 109
   :||||| :||||| :|||
1114 TTATAGTGGTAAAGCTAGCTCTTAAGGGCTTTTGAATTTCCAGCAAGC 1163
   :||||| :||||| :|||
109 lPheAsnAspProAspHisSerValThrGlyAsnAsnLeuGluPheTyrG 126
   :||||| :||||| :|||
1164 CTTTACAAATCAACACCACTAACATCAACAGCTTCTCTTT...A 1210
   :||||| :||||| :|||
126 lyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValAla 142
   :||||| :||||| :|||
1211 ATAACGCCACTTTTAAACAAT.....ACGGGTAAATCACTATA 1248
   :||||| :||||| :|||
143 AspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeuAl 159
   :||||| :||||| :|||
1249 GAAAAA.....GATCGAGCTTTTAAATAAC..... 1272
   :||||| :||||| :|||
159 aProGluGluAlaThrArgGlnValMetGlyTyrThrGluGluGluV 176
   :||||| :||||| :|||
1272 ..... 1272
176 alProThrLeuValAspLeuValAspGluPheThrPheAsnSerTrp 192
   :||||| :||||| :|||
1273 .....ACGACATTCACACT... 1287
   :||||| :||||| :|||
193 PheSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeuAl 209
   :||||| :||||| :|||
1288 .....TCTGTGTATACAAACAAACATGAGTGTACCGG 1319
   :||||| :||||| :|||
209 aglyThrAlaAlaGlyHisGlyLysAsnAsp.....AspAspPheL 223
   :||||| :||||| :|||
1320 TGGCGTTACTTTAAGC...GGTAAATGACTTCAAAATGGCTCAACCC 1366
   :||||| :||||| :|||
223 euAsnTyrGlyLeuSerSerLys..... 230
   :||||| :||||| :|||
1367 TTGATTTTGGG...AGTTCTAAATCACTCTCGTCAAGGACGACTTTC 1413
   :||||| :||||| :|||
231 SerIlePheGluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAs 247
   :||||| :||||| :|||
1414 AACCTCACAGCTTTAGCAGTGAAGAAGCGCTAACGATTTTAAATCTAG 1463
   :||||| :||||| :|||
247 pGlyThrAsnGlyGluPheGluProAspSerLeuPhePheThrTyrVala 264
   :||||| :||||| :|||
1464 CGGTGGG.....ATCACTTATAGTA 1483
   :||||| :||||| :|||
264 sn.....GlnThrSer 267
   :||||| :||||| :|||
1484 ACCTTTAAACCATGCAATCAACGGCTTGACAAGTGCCTTAAACACGAAC 1533
   :||||| :||||| :|||
268 ArgSerAsnValValProValGluAsnPhePheGlnAsp..... 280
   :||||| :||||| :|||
1534 GAAAGCTTTCAAAATCCG...CAAAGTTTCGCTCAAGGTTTGTGGGATAT 1580
   :||||| :||||| :|||
281 ...AlaTyrLeuGlyValLeuProLysPheSerTyrIleAsn..... 293
   :||||| :||||| :|||
1581 AATCACTTACAATGGGTTTACCGGCGAGCTTTTGAATGAAACGCTGCAA 1630
   :||||| :||||| :|||
294 .....ProSerCysCysGlyThrAsnThrAsn 302
   :||||| :||||| :|||
1631 CATCTAAACCCACTGACTCTCGCCCTCTTAATCTCTCAAACTCTACG 1680
   :||||| :||||| :|||
303 SerMetHisProThrGluAsnValSerTyrGlyGluValPheValLysG 319
   :||||| :||||| :|||
1681 CAAGTCTATCAAGTGGGT...TACAAATAGGGGATACTATCTACAAACT 1727
   :||||| :||||| :|||
319 n.....IleTyrAspAlaIleArgGlnG 327

```

```

1728 GCAGAAACTTTGAGCACACATTCATTATTTAGAGAGCG 1777
   :||||| :||||| :|||
327 lyProGlnTrpAspLysThrLeuPheIleThrTyrAspGluThrGly 343
   :||||| :||||| :|||
1778 GG.....ACTTAC..... 1785
   :||||| :||||| :|||
344 GlyPheTyrAspHisValProProLeuAlaValArgProAspAsnLe 360
   :||||| :||||| :|||
1786 .....ACGCCACCCCTGTCAAT..... 1803
   :||||| :||||| :|||
360 uThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheA 377
   :||||| :||||| :|||
1804 .....AACGGCTCCAAATTTGACTTATTCGGCTT 1831
   :||||| :||||| :|||
377 spArg...LeuGlyGlyArgMetProThrTrpValIleSerProTyrSer 392
   :||||| :||||| :|||
1832 CAAATATATCAATGCTGACATGCCT...TGG.....TATGAC 1866
   :||||| :||||| :|||
393 LysLysGlyTyrIle 397
   :||||| :||||| :|||
1867 CATAAATATTAATC 1881
   :||||| :||||| :|||

```

seq\_name: /cgn2\_2/cgdata/geneseq/NA2000.DAT.Z49213

seq\_documentation\_block:

ID Z49213 standard: DNA; 2264 BP.

XX AC Z49213;

XX DT 15-MAR-2000 (first entry)

XX DE Adeno associated virus AAV5 capsid protein VP2 DNA.

XX KW Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR; promoter; Rep protein; capsid protein; regulation; transcription; replication; chromosomal integration; tissue tropism; cellular receptor; gene therapy; neutralising antibody; erythroid progenitor cell; transduction; cancer; genetic disease; VP2; ss.

XX OS Adeno associated virus type 5.

XX Key Location/Qualifiers

FT CDS 470..2236

FT /tag= a

FT /product= "AAV5 capsid protein VP2"

FT /note= "The translation initiation codon in this sequence is ACG, which encodes Thr; this unusual start codon is specifically mentioned in the specification"

XX W09961601-A2.

XX PD 02-DEC-1999.

XX PF 28-MAY-1999; 99NO-US11958.

XX PR 28-MAY-1998; 98US-0087029.

XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX PI Chlorini JA, Kotin RM;

XX DR WPT: 2000-062707/05.

XX DR P-PSDB; 158161.

XX PT Adeno-associated virus 5 based vectors and particles, useful for gene therapy

XX PS Claim 31; Page 83; 91pp; English.

XX CC This sequence represents DNA encoding the VP2 capsid protein of adeno associated virus type 5 (AAV5). The invention relates



FT CDS 638...2236  
 FT /\*tag= a  
 XX /product= "AAV5 capsid protein vp3"  
 XX  
 PN W09961601-A2.  
 XX  
 PD 02-DEC-1999.  
 XX  
 PF 28-MAY-1999; 99WO-US11958.  
 XX  
 PR 28-MAY-1998; 98US-0087029.  
 XX  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA  
 XX Chiorini JA, Kotin RM;  
 XX WPI: 2000-062707/05.  
 DR P-PSDB: Y58162.  
 XX  
 XX Adeno-associated virus 5 based vectors and particles, useful for gene  
 PT therapy  
 PT  
 PT  
 PS Claim 32; Page 83-84; 91pp; English.  
 XX  
 CC This sequence represents DNA encoding the VP3 capsid protein of  
 CC adeno associated virus type 5 (AAV5). The invention relates  
 CC to vectors comprising a pair of AAV5 inverted terminal repeats  
 CC (ITRs) with a promoter between the ITRs. The vector may comprise  
 CC the viral genome, or subregions thereof, including sequences  
 CC encoding Rep proteins and capsid proteins, and is encapsidated  
 CC in an AAV5 particle. The non-structural Rep proteins Rep40  
 CC (Y58613), Rep52 (Y58166), Rep68 (Y58164) and Rep78 (Y58159) are  
 CC involved in regulation of replication and transcription, in addition to  
 CC the production of progeny genomes. Rep68 and Rep78 are also associated  
 CC with the stable integration of the viral genome into human chromosomes.  
 CC The three types of capsid protein VP1 (Y58160), VP2 (Y58161) and VP3  
 CC (Y58162) assemble to form an icosahedral capsid, and differ from each  
 CC other by the use of alternative splicing and an unusual translation  
 CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2  
 CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are  
 CC likely to utilise distinct cellular receptors and are serologically  
 CC distinct. In a gene therapy application, therefore, AAV5 would allow for  
 CC transduction of a patient who already possess neutralising antibodies  
 CC either as a result of natural immunological defence or from prior  
 CC exposure to AAV2 vectors. The vectors may be useful for transducing  
 CC erythroid progenitor cells or cells lacking heparin sulphate  
 CC proteoglycans, which is very inefficient with AAV2-based vectors.  
 CC The vectors may also be useful for transducing cells with a nucleic  
 CC acid of interest in order to produce cell lines that could be used to  
 CC screen for agents that interact with the gene product of the nucleic  
 CC acid of interest. In addition to transduction of other cell types,  
 CC transduction of erythroid cells would be useful or the treatment of  
 CC cancer and genetic diseases which can be corrected by bone marrow  
 CC transplants using matched donors.  
 XX  
 SQ Sequence 2264 BP; 575 A; 722 C; 595 G; 372 T; 0 other;

alignment\_scores:  
 Quality: 107.00 Length: 346  
 Ratio: 0.738 Gaps: 16  
 Percent Similarity: 41.908 Percent Identity: 19.075  
 alignment\_block:  
 US-09-426-072-2 x Z49214 ..  
 Align seg 1/1 to: Z49214 from: 1 to: 2264  
 118 ThrGlyAsnAsnLeuGlu.....PheTy 125  
 1253 ACGGCAACAACCTTTGAGTTTACCTACCACTTTGAGAGGTCCTTCCA 1302  
 125 rGlyThrThrProAsnAsnGlyAlaIleAlaSerGlyLysValVa 142

1303 CTCGAGCTTCGCTCCAGTCACAGCTGTCACAGCTGGCCACCGCTGG 1352  
 142 IaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeu 158  
 1353 TGGACCACTACTTGTACCGCTTCGTGAGCACAATAAC.....ACT 1393  
 159 AlaProGluGluAlaThrArgGlnValMetGlyTyrThrGluGluG1 175  
 1394 GCGGAGTCCAGTTCACAGAACCTGCCGGGAGATACGCC..... 1435  
 175 uValProThrLeuValAspLeuValAspGluPheThrPheAsnSerT 192  
 1436 .....AACACTACAAAAC 1451  
 192 rpPheSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeu 208  
 1452 GGTTC.....CCGGGCCCATGGCGCAACCCAG..... 1480  
 209 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 225  
 1481 ..GGCTGGAACTCGGGCTCCGGGGTCAACCGC..... 1510  
 225 rGlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLysGlyVals 242  
 1511 .....GCCAGTGTGAGCGCTTCGCCAGCAGCCATAGG..... 1543  
 242 erTTPLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAspSerLeu 258  
 1544 .....ATGGAGCTCGAGGGCGGAGTTACCGAGTGCCTCCGCGCAGCGAAC 1588  
 259 PhePheThrTyrValAsnGlnThrSerArgSerAsnValValProValG1 275  
 1589 GCGATGACC.....AACAACTCCAGGGCAGCAACACCTATGCCCTGGA 1632  
 275 uAsnPhPhe.....GlnAspAlaTyrLeuGlyValLeuProL 288  
 1633 GAACCATATGATCTTCAACAGCGCGGCGGCAACCGCGGCACCGCC. 1681  
 288 ysPheSerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMet 304  
 1682 .....ACGTACCTCGAGGGCAACATGCTCATCACCAGCGAGGAGACG 1726  
 305 HisProThrGlyAsnValSerTyr..... 312  
 1727 CAGCGGTGAACCGCTGGCGGTACACGTCGCGGGCAGATGGCCACCA 1776  
 312 ..... 312  
 1777 CAACGAGAGCTCCACCACTGCCCGCGGCGGCGGCGGCGGCGGCGG 1826  
 313 .....GlyGluValPheValLysGlnIleTyrAspAlaIleArg 325  
 1827 AAATCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1870  
 326 GlnGlyProGlnThrAspLysThrLeuLeuPheIleThrTyrAspGlu 342  
 1871 CAAGGACCCATCTGGGCGCAAG.....ATCCCGAGAGAC 1902  
 342 rGlyGlyPheTyr.....AspH 348  
 1903 GGGGGCGCACTTTCACCCCTCTCCGGCCATGGCGGATTCGGACTCAAC 1952  
 348 lsValProProLeuAlaValArg.....ProAspAsnLeu 360  
 1953 ACCCAGCGCCATGATGCTCATCAAGAACAACGCTGTGCCGGGAATATC 2002  
 361 ThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAs 377  
 2003 ACC..... 2005  
 377 parGLeuGlyGlyArgMetProThrTrpValIleSerProTyrSerLys 394

```

2006 .....ACCTTCGCGAGTGGCCGTCA 2027
394 ySGlyTYrileGluInTyRgLyThrAspProValThr 406
      ::::||||| |||||:::|||||
2028 GCAGCTTCATCACCAGTACAGCACCGCGAGGTACC 2065

seq_name: /cgn2_2/gcgdata/geneseq/NA2000.DAT:Z49212

seq_documentation_block:
ID   Z49212 standard; DNA; 2307 BP.
XX
AC   Z49212:
XX
DT   15-MAR-2000 (first entry)
XX
DE   Adeno associated virus AAV5 capsid protein VP1 DNA.
XX
KW   Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;
KW   promoter; Rep protein; capsid protein; regulation; transcription;
KW   replication; chromosomal integration; tissue tropism; cellular receptor;
KW   gene therapy; neutralising antibody; erythroid progenitor cell;
KW   transduction; cancer; genetic disease; VP1; ss.
XX
OS   Adeno associated virus type 5.
XX
FH   Key Location/Qualifiers
FT   CDS 105..2279
FT     /tag= a
FT     /product= "AAV5 capsid protein VP1"
XX
XX   WO9961601-A2.
XX
XX   02-DEC-1999.
XX
XX   28-MAY-1999; 99WO-US11958.
XX
XX   28-MAY-1998; 98US-0087029.
XX
XX   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX   Chlorini JA, Kotin RM;
XX
XX   WPI: 2000-062707/05.
XX
XX   P-PSDB; Y58160.
XX
PT   Adeno-associated virus 5 based vectors and particles, useful for gene
PT   therapy.
XX
PS   Claim 30; Page 82; 91pp; English.
XX
XX   This sequence represents DNA encoding the VP1 capsid protein of
XX   adeno associated virus type 5 (AAV5). The invention relates
XX   to vectors comprising a pair of AAV5 inverted terminal repeats
XX   (ITRs) with a promoter between the ITRs. The vector may comprise
XX   the viral genome, or subregions thereof, including sequences
XX   encoding Rep proteins and capsid proteins, and is encapsidated
XX   in an AAV5 particle. The non-structural Rep proteins Rep40
XX   (Y58161), Rep52 (Y58168), Rep68 (Y58164) and Rep78 (Y58159) are
XX   involved in regulation of replication and transcription, in addition to
XX   the production of progeny genomes. Rep68 and Rep78 are also associated
XX   with the stable integration of the viral genome into human chromosomes.
XX   The three types of capsid protein VP1 (Y58160), VP2 (Y58161) and VP3
XX   (Y58162) assemble to form an icosahedral capsid, and differ from each
XX   other by the use of alternative splicing and an unusual translation
XX   initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2
XX   capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are
XX   likely to utilise distinct cellular receptors and are serologically
XX   distinct. In a gene therapy application, therefore, AAV5 would allow for
XX   transduction of a patient who already possess neutralising antibodies
XX   either as a result of natural immunological defence or from prior
XX   exposure to AAV2 vectors. The vectors may be useful for transducing
XX   erythroid progenitor cells or cells lacking heparin sulphate
XX   proteoglycans, which is very inefficient with AAV2-based vectors.

```

```

CC   The vectors may also be useful for transducing cells with a nucleic
CC   acid of interest in order to produce cell lines that could be used to
CC   screen for agents that interact with the gene product of the nucleic
CC   acid of interest. In addition to transduction of other cell types,
CC   transduction of erythroid cells would be useful or the treatment of
CC   cancer and genetic diseases which can be corrected by bone marrow
CC   transplants using matched donors.
XX
SQ   Sequence 2307 BP; 587 A; 731 C; 603 G; 386 T; 0 other;

```

```

alignment_scores:
  Quality: 107.00      Length: 346
  Ratio: 0.738        Gaps: 16
  Percent Similarity: 41.908      Percent Identity: 19.075

alignment_block:
US-09-426-072-2 x Z49212 ..
Align seg 1/1 to: Z49212 from: 1 to: 2307

118 ThrGlyAsnAsnLeuGlu.....PheTy 125
|||||
1296 ACGGCGACACACTTTCAGTTTACCTACACTTTGAGGAGGTCCCTTCCA 1345
|||||
125 rGlyThrTyThrProAsnAsnGlyAlaIleAlaSerGlyLysValValA 142
|||||
1346 CTCCAGCTTCGCTCCAGTCAGAACCTGTTCAAGCTGGCCCAACCGCTGG 1395
|||||
142 IaAspGInSerGlyPheLeuAsnAlaGlnLeuAsnAspTyThrProLysLeu 158
|||||
1396 TGGACCACTACTTGTACCGCTTCGTGAGCACAATAAC.....ACT 1436
|||||
159 AlaProGluGluAlaThrArgGlnValMetGlyTyThrGluGluGlu 175
|||||
1437 GCGGAGTCCAGTTCACAAAGAACCTGGCGGGAGATACGCC..... 1478
|||||
175 uValProThrLeuValAspLeuValAspGluPheThrThrPheAsnSerT 192
|||||
1479 .....AACACCTACAAAACACT 1494
|||||
192 rPheSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeu 208
|||||
1495 GGTTC.....CCGGGGCCCATGCGCCGCAACCCAG..... 1523
|||||
209 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 225
|||||
1524 ...GGCTGGAACCTGGGCTCCGGGTCAACCGC..... 1553
|||||
225 rGlyIleSerSerIlePheGluAlaAlaAsnGluLysGlyValS 242
|||||
1554 .....CCCAAGTGTAGCGCTTCGCCAGCACAATAGG..... 1586
|||||
242 eTrpLeuAsnTyAspGlyThrAsnGlyGluPheGluProAspSerLeu 258
|||||
1587 .....ATGGAGCTCGAGGGCGGAGTTACCAAGGTGCGCCGCGCGCAAC 1631
|||||
259 PhePheThrTyValAsnGlnThrSerArgSerAsnValValProValG 275
|||||
1632 GGCATGAC.....AACACCTCCAGGCGGACACACCTATGCTCCCTGGA 1675
|||||
275 uAsnPhePhe.....GlnAspAlaTyLeuGlyValLeuProL 288
|||||
1676 GAACACTATGATCTTCAACAGCCAGCCGCGCAACCGGGCACCACCGCC 1724
|||||
288 yPheSerTyIleAsnProSerCysGlyThrAsnThrAsnSerMet 304
|||||
1725 .....ACGTACTCGAGGGCACAATGCTCATCACCAGGAGAGACAG 1769
|||||
305 HisProThrGlyAsnValSerTy..... 312
|||||
1770 CAGCGGTTGAACCGCGTGGCGTACACGTCGCGGGGCGAGATGCCACCAA 1819
|||||

```

```

312 ..... 312
1820 CAACAGAGCTCCACCACTGCCCGCCGACCGGACGTACACACTCCAGG 1869
313 .....GlyGluValPheValLysGlnIleTyrAspAlaIleArg 325
1870 AATCGTGGCCGCGAGGTGTGGATGGAGAGG.....GACGTGTACCTC 1913
326 GlnGlyProGlnTTPAspLysThrLeuLeuPheIleThrTyrAspGlu 342
1914 CAAGGACCCATCTGGGCCAAG.....ATCCAGAGAGAC 1945
342 rGlyGlyPheTyr.....AspH 348
1946 GGGGGCGCATTTACCCCTCCGCCCATGGCGGATTCGGACTCAAAAC 1995
348 isValProProLeuAlaValArg.....ProAspAsnLeu 360
1996 ACCACCGCCCATGATGCTCATCAAGAGACAGCGCTGTGCCCGGAATATC 2045
361 ThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAs 377
2046 ACC..... 2048
377 pArgLeuGlyArgMetProThrTrpValIleSerProTyrSerLysL 394
2049 .....AGCTTCTCGGACGTGCCCGTCA 2070
394 ysGlyTyrIleGluLysThrGlyThrAspProValThr 406
2071 GCAGCTTCATACCCAGTACAGACGCCGGGCGAGGTACCC 2108
seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1992.DAT.Q31411
seq_documentation_block:
ID Q31411 standard; DNA; 3771 BP.
XX AC Q31411;
XX DT 20-APR-1993 (first entry)
XX DE Bacillus thuringiensis gene 33F2.
XX KW Toxin protein; ant; ds.
XX OS Bacillus thuringiensis isolate 33F2.
XX FH Key Location/Qualifiers
XX CDS 1..3771
FT /*tag= a
FT /note= "protein 33F2"
FT 4..24
FT /*tag= b
FT /note= "oligonucleotide hybridisation probe a"
FT 13..33
FT /*tag= c
FT /note= "oligonucleotide hybridisation probe b"
XX
XX PN W09220802-A.
XX
XX PD 26-NOV-1992.
XX
XX PF 22-MAY-1992; 92WO-US04316.
XX
XX PR 22-MAY-1991; 91US-0703977.
XX PR 25-NOV-1991; 91US-0797645.
XX PR 12-MAY-1992; 92EP-0304228.
XX
XX PA (MYCO ) MYCOGEN CORP.
XX PA Kennedy MK, Meier H, Payne JM, Randall JB, Uick HJ;
XX PI WPI; 1992-415780/50.
XX DR

```

```

DR P-PSDB; R29028.
XX Toxin proteins isolated from Bacillus thuringiensis - for controlling
PT ants. e.g. fire, carpenter, argentine and pharaoh ants
XX
XX Disclosure; Page 44; 71pp; English..
XX
CC Bacillus thuringiensis gene 33F2 codes for the toxin protein 33F2.
CC This is useful as a method of biological control of ants, e.g. fire
CC ants, carpenter ants, argentine ants and pharaoh ants, as an
CC alternative to chemical insecticides.
XX
XX Sequence 3771 BP; 1417 A; 615 C; 590 G; 1149 T; 0 other;
SQ
alignment_scores:
Quality: 107.00 Length: 491
Ratio: 0.493 Gaps: 22
Percent Similarity: 44.196 Percent Identity: 19.552
alignment_block:
US-09-426-072-2 x Q31411 ..
Align seg 1/1 to: Q31411 from: 1 to: 3771
6 LeuLeuThrGlyLeuGlyLeuAlaSerLeuGly...LeuAlaSerPr 21
220 GITTATCTGTATTAACTTTAATCTCTCAAGTTGGTACTGTGGCTCTGC 269
21 oValThrSerGluTyrThrSerValArgGluAlaProPheGlyTyrLysp 38
270 AGCAAGTACTATTGTAAGTTTATTGGCCTTAAATATTGGAGATAAAC 319
38 roGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIle 54
320 CAATGCAAAAATATATTGAGAGCTCAAGCTCAAAATTCAGACATTA 369
55 ValTrpLeuIleLeuGluAsn.....ArgSerPh 64
370 ATTCAACAGATATACAACTATCAAGTGCATTAATCAAAAAAAT 419
64 eAspAsnIleLeuGlyValArgArgGlnGlyLeuAspAsnProIleA 81
420 TGACAGTCTTCAGAAACAAATTAATCTATATACAGTA.....GCTATAG 463
81 snAsnGlyProPheCysAsnTyrLysAsnAlaSerAspProSerSerCly 97
464 ATACATGATTACGTACACGAAAA..... 489
98 LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAsnAsp..... 112
490 .....ACGCAACTCGAAAAATCTAAATTTCTATACTTACCTCAGATAT 530
113 .....ProAspHisSerValThrGlyAsnAsnLeuGluPheT 125
531 CTCCTATTATTCCAGAGGATATGAACCTGGA...GGTTTACCTTATT 577
125 yrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysVal 141
578 ATGCTATGGTTGCTAATGCTCATATATTATTGTTAAGAGAGCGCTATAGT 627
142 ...AlaAspGlnSerGlyPheLeuAsnAlaClnLeuAsnAspTyrProLy 157
628 AATGCAGAGAAATAGGCTTTAGTGTAAAGAGTAGACACACATAAAAA 677
157 sLeuAlaProGluAlaThrArgGlnValMetGlyTyrTyrThrGluG 174
678 ATAT.....ATCAAAATGACAATACACATCATCTAGTGAAG 712
174 luGluValProThrLeuValAspLeuValAspGluPhe..... 186
713 CAGTAATAAAGCAATCTTAAATGCACTTGACAAATTTAAGAGTTTAGAT 762

```





```

270 AGCAAGTACTATTGTAAGTTTATTTGGCCATAAATATTGGAGATAAAC 319
      :::::::::::   ::::   |||||   ||||
38  roGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIle 54
      ||::||::||::   ::||::||::||::||::||::||::||::||::
320 CAATCGAANAATATATTGAGAGCTCAAGCTCAAAATGAGCATTG 369
      ||::||::||::   ::||::||::||::||::||::||::||::||::
55  ValTrpLeuIleLeuGluAsn.....ArgSerPh 64
      ::::   ::::   |||   ::::|||
370 ATTCAACAAGATATAACAACACTATCAAGATGCAATTAATCAAAAAAAT 419
      |||||   ::::   ::|||   ::|||
64  eAspAsnIleLeuGlyValAlaArgGlnGlyLeuAspAsnProIleA 81
      |||||   ::::   ::|||   ::|||
420 TGACAGCTCTGAGAAACAATTAATCTATATACAGTA.....GCTATAG 463
      |||||   ::::   ::|||   ::|||
81  snAsnGlyProPheCysAsnTyrLysAsnAlaSerAspProSerSerGly 97
      ::||::||::   ::::   |||   ::|||
464 ATAACAATGATTAGCTACACAGCAAA..... 489
      |||||   ::::   ::|||   ::|||
98  LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAsnAsp..... 112
      |||||   ::::   ::|||   ::|||
490 .....ACGCAACTCGAAATCTAAATCTATCTTACCTCAGATAT 530
      |||||   ::::   ::|||   ::|||
113 .....ProAspHisSerValThrGlyAsnAsnLeuGluPheT 125
      |||||   ::::   ::|||   ::|||
531 CTCATATTTATCCAGAGGATATGAAACTGGA...GTTTACCTTAT 577
      |||||   ::::   ::|||   ::|||
125 yrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141
      |||||   ::::   ::|||   ::|||
578 ATGCTATGTTGCTAATGCTCATATATTATTGTTAAGAGACGCTATAGT 627
      |||||   ::::   ::|||   ::|||
142 ...AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProly 157
      ||::||::||::   ::||::||::||::||::||::||::||::||::
628 AATGCAGAGAATAGCTTTAGTGATAAAGAAGTAGACACACATAAAAA 677
      ||::||::||::   ::||::||::||::||::||::||::||::||::
157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGluG 174
      ||::||::||::   ::||::||::||::||::||::||::||::||::
678 ATAT.....ATCAAAATGACAATACACAATCATACTGAAG 712
      ||::||::||::   ::||::||::||::||::||::||::||::||::
174 LuGluValProThrLeuValAspLeuValAspGluPhe..... 186
      ::::   ::::   ::|||   ::|||
713 CAGTAATAAGCAATCTTAAATGGACTTGACAAATTTAAGAGTTTAGAT 762
      ||::||::||::   ::||::||::||::||::||::||::||::||::
187 ...ThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn... 201
      ||::||::||::   ::||::||::||::||::||::||::||::||::
763 GTAATAGCTATATAAAGCAAAATATATTAAGGTATGACAGAAAT 812
      ||::||::||::   ::||::||::||::||::||::||::||::||::
202 .ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217
      |||||   ::::   |||   |||||
813 GGTCTTGATCTAGTCTGCTATGCGCACTTTGTCGATCCAGATCAITATC 862
      |||||   ::::   |||   |||||
217 ysAsnAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePhe 233
      ::::||::||::   ::||::||::||::||::||::||::||::||::
863 AAAAGAGATGAAATGAAATTTACAAGAAGCTATTTCTCTCAATTTAC 912
      ||::||::||::   ::||::||::||::||::||::||::||::||::
234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAs 250
      ::::   ::::   |||   ::|||
913 CAACCTGTACCTAAA.....AACATGCAAAATACCTC 944
      ||::||::||::   ::||::||::||::||::||::||::||::||::
250 nGlyGluPheGluProAspSerLeuPhe..... 259
      ::::||::||::   ::||::||::||::||::||::||::||::||::
945 TAGCTCTATTGTACTAGCTAGCGATCTATTTCACTATCAAGGAGATCTGTAA 994
      ||::||::||::   ::||::||::||::||::||::||::||::||::
260 .....Phe 260
      ||::||::||::   ::||::||::||::||::||::||::||::||::
995 AATTAGAAATTTCTACAGAACGGACACGATGCTTGCAGAAAATTTT 1044
      ||::||::||::   ::||::||::||::||::||::||::||::||::
261 ThrTyrValAsnGlnThrSerArgSerAsnValValProValGluAsnPh 277
      |||||   ::::   |||   |||||
1045 ACTGGTATTTCGAAACACATCTTACAAATCGCCTAATACTCATGAACATA 1094
      |||||   ::::   |||   |||||
277 ePheGlnAspAlaTyrLeuGlyValLeuProLysPheSerTyrIleAsnPh 294
      ::::   ::::   |||   |||||

```

```

1095 CCATGTAGAT.....TTTAGTTAT..... 1113
294 roSerCysCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 310
      |||||   ::::   |||||
1114 .....ATATCCCAATCT.....AGTGGTAATATT 1137
      |||||   ::::   |||||
311 SerTyrGly..... 313
      |||   |||
1138 TCAAGAGGCTCTTCAATCCGATTCCAAATTCATTTAATAATATCCATTAT 1187
      |||||   ::::   |||||
314 .....GluValPheValLysGlnIleTyrAsp..... 322
      ::::   ::::   |||   |||
1188 TTCAACTTGTATTAGAAATTCATTTTATAGGCANATAGCGGATCTTCTG 1237
      ::::   ::::   |||   |||
323 .....AlaIleArgGlnGly 327
      |||||   ::::   |||||
1238 TTTTAGTTTAATTTTAAAGATGGCACTCAAGGGTATGCATTTGCCCAAGCA 1287
      |||||   ::::   |||||
328 ProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyG 344
      |||||   ::::   |||||
1288 CCA.....ACAGGAGG 1298
      |||||   ::::   |||||
344 yPheTyrAspHis.....ValProProLeuAlaValArgProAspA 359
      ::||::||::||::   ::||::||::||::||::||::||::||::||::
1299 TGCCTGGACCATTTCTTTATTGTAATCTGTATGGTCCGCCAGAAAGGGCATA 1348
      |||||   ::::   |||||
359 snLeuThrTyrThrGluThrAlaLysAsnGly..... 369
      ::||::||::||::   ::||::||::||::||::||::||::||::||::
1349 AATTAACATATATTTATCTTCTCCAGGTGATACATTAAGAGATTTTCATC 1398
      |||||   ::::   |||||
370 GlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 386
      ::||::||::||::   ::||::||::||::||::||::||::||::||::
1399 AATGTATATACTCTTAAAGTACTCCAATATAAATGAATCAACAGATA 1448
      |||||   ::::   |||||
386 pValIleSerProTyr...SerLysLysGlyTyrIleGluGlnTyrGlyT 402
      ||::||::||::   ::||::||::||::||::||::||::||::||::
1449 AAAAATCAAGGCTTCTCTCGGAAGAGGATATATCAAAATCAAGGA 1498
      |||||   ::::   |||||
402 hrAspProValThrGlyLysPro 409
      |||||   ::::   |||||
1499 TCATGAAATATTACGGTAAACCA 1521

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/NA1992.DAT.Q30937

seq\_documentation\_block:

ID Q30937 standard; DNA; 3771 BP.

XX Q30937;

XX DT 30-MAR-1993 (first entry)

XX BT toxin 33F2.

XX nematode worms; nematocides; nematocidal toxin; agriculture; plants;

XX crops; pests; CryV proteins.

XX OS Bacillus Thuringiensis.

XX Key Location/Qualifiers

XX CDS 1..3771

XX /\*tag= a

XX PN W09219739-A.

XX PD 12-NOV-1992.

XX PF 01-MAY-1992; 92WO-US03624.

XX PR 03-MAY-1991; 91US-0693018.

XX PR 31-JAN-1992; 92US-0830050.

XX PR 23-APR-1992; 92US-0871510.

XX PA (MYCO ) MYCOGEN CORP.

XX  
PI Foncerrada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;  
XX  
XX WPI: 1992-398866/48.  
XX DR p-PSDB; R28811.  
XX  
XX New genes and toxins against nematodes - obtd. from *Bacillus*  
XX PT *Thuringiensis* isolates with nematocidal activity  
XX  
XX Claim 1(b) : Page 53; 77pp; English.  
XX  
XX This sequence encodes the *Bacillus thuringiensis* delta-endotoxin  
XX CC having nematocidal activity. A library was constructed in  
XX CC lambdaDEMI1 using Bt strain P33F2 protoplast total DNA digested  
XX CC with EcoRI and size fractionated by electrophoresis  
XX CC (4.3-6.6 used). Fragments were ligated into pHTBlueII (an E. coli/BT  
XX CC shuttle vector comprised of pBluescript S/K and a replication origin  
XX CC from a resident Bt plasmid). This was used to transform NM522  
XX CC cells, and colonies screened with a probe PCR amplified using primers  
XX CC Q30934..55. The desired plasmid pMYC2316 contained an approx.  
XX CC 5.85Kbp EcoRI insert, the toxin gene (33F2a) residing on this  
XX CC fragment being novel compared to sequences encoding other  
XX CC nematocidal proteins.  
XX  
XX SQ Sequence 3771 BP; 1419 A; 614 C; 590 G; 1148 T; 0 other;

alignment\_scores:  
Quality: 107.00 Length: 491  
Ratio: 0.493 Gaps: 22  
Percent Similarity: 44.196 Percent Identity: 19.552

alignment\_block:

US-09-426-072-2 x Q30937 ..

Align seg 1/1 to: Q30937 from: 1 to: 3771

6 LeuLeuThrGlyLeuGlyLeuLeuAlaSerLeuGly...LeuAlaSerPr 21  
220 GTTTTATCTCTATTAACTTTAAATTCCTGAAGTTGGTACTGTGGCTCTGC 269  
21 oValTrpSerGluTyrThrSerValArgGluAlaProPheGlyTyrLysP 38  
270 AGCAGTACTATTGTAAGTTTATTTGGCTAAATATTTGGAGATAAAC 319  
38 toGlySerLysGluSerLeuGlyLeuAlaSerLysLysValGluAlaSer 54  
320 CAAATGCAAAAATATATTGAGAGCTCAAGCTCAAAATTTGAAGCATT 369  
55 ValTrpLeuLeuLeuGluAla...ArgSerPh 64  
370 ATTCAACAAGATATACAACTATCAAGATGCAATTAATCAAAAAAATT 419  
64 eAspAlaLeuGlyGlyValArgGlnGlyLeuAlaSerProLeuAla 81  
420 TGACAGTCTTCAGAAAACAATTAATCTATATACAGTA.....GCTATAG 463  
81 snAsnGlyProPheCysAsnTyrLysAlaSerAlaSerAspProSerSerGly 97  
464 ATACATGATTACGTACACGCAAA..... 489  
98 LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAlaSerAsp 112  
490 .....ACGCAACTCCAAAATCTAAATTTCTATCTACTTACCTCAGATAT 530  
113 .....ProAspHisSerValThrGlyAlaSerLeuGluPheT 125  
531 CTCCATATTTATTCAGAGAGGATATGAAGTGA...GGTTACCTTATT 577  
125 yGlyThrTyrProAlaSerAlaLeuAlaSerGlyLysValVal 141  
578 ATGCTATGGTTGCTATATATTATTTGTTAGAGAGCTATAGTT 627

142 ...AlaAspGlnSerGlyPheLeuAlaGlnLeuAlaSerProLys 157  
628 AATGCAGAGAAATTTAGCTTTAGTGATAAAGAAAGTAGACACACATATAA 677  
157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrThrGluG 174  
678 ATAT.....ATCAAAATGACAATACACAATCATACTACTGAAG 712  
174 luGluValProThrLeuValAspLeuValAspGluPhe..... 186  
713 CAGTAATAAAGCATCTTAAATGGACTTGACAAATTTAAGAGTTTAGAT 762  
187 ...ThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn... 201  
763 GTAATAGGTATATAAATAAAGCAATATATTAAGGTATGACAGAAAT 812  
202 .ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217  
813 GGTTCCTGTAGTCTGTCTCTATGGCAACTTTTCGATCCAGATCATATC 862  
217 ysAsnAspAspPheLeuAlaSerGlyLysSerLysSerLysPhe 233  
863 AAAAGAAGTAGAAATTCGAATTTACAGAAGCAATTTCTCTCCAAATTTAC 912  
234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAlaSerGlyThrAs 250  
913 CAACCTGTACCTAAA.....AACATGCAAAATACCTC 944  
250 nGlyGluPheGluProAspSerLeuPhe..... 259  
945 TAGCTCTATTGTACCTAGCATCTATTTCACCTATCAAGGAGATCTGTAA 994  
260 .....Phe 260  
995 AATTAGAAATTTCTACAGAAGCGACACAGATGGTCTTGCAGAAATTTT 1044  
261 ThrTyrValAsnGlnThrSerArgSerAsnValProValGluAlaSer 277  
1045 ACTGGTATTGCAACACATCTTCAAAATCGCTTAATACTCATGAACATA 1094  
277 ePheGlnAlaLysLeuGlyValLeuProLysPheSerTyrIleAsnP 294  
1095 CCATGTAGAT.....TTAGTTAT..... 1113  
294 roSerCysCysGlyThrAsnThrAsnSerMethIleProThrGlyAla 310  
1114 .....AATACCAATCT.....AGTGGTATATT 1137  
311 SerTyrGly..... 313  
1138 TCAAGAGGCTCTTCAAAATCCGATTCGAATTTAATAATCCCATTTAT 1187  
314 .....GluValPheValLysGlnIleTyrAsp..... 322  
1188 TTCACCTGTTAGAAATTCATTTTAAAGCAATAGCGGGATCTCTG 1237  
323 .....AlaIleArgGlnGly 327  
1238 TTTTAGTTAATTTTAAAGATGGCACTCAAGGGTATGCATTTGCCCAAGCA 1287  
328 ProGlnIleProLysThrLeuPheIleThrTyrAspGluThrGlyG 344  
1288 CCA.....ACAGGAGG 1298  
344 yPheTyrAspHis.....ValProProLeuAlaValArgProAspA 359  
1299 TGCCTGGGACCATTCTTTTATTGAATCTGTGGTGGCCCAAGAGGCATA 1348  
359 snLeuThrTyrThrGluThrAlaLysAsnGly..... 369  
1349 AATTAACTATATTATTTACTTCTCCAGGTATCATTAAGAGATTTTCATC 1398  
370 GlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 386

```

... |||||
1399 AATGTTATATCTCTTATAAGTACTCCAACTATTAATGAACATCAACAGA 1448
386 pValIleSerProTyr...SerLysLysGlyTyrIleGluGlnTyrGlyT 402
|||...: : : : : |||||
1449 AAAATCAAGGCTTTCTCGCGGAAAAGGATATATCAAAAATCAAGGGA 1498
402 hrAspProValThrGlyLysPro 409
1499 TCATGAAATATTACGGTAAACCA 1521

seq_name: /cgn2_2/gcgdata/geneseq/geneseq/NA1993.DAT:Q51689
seq_documentation_block:
ID_Q51689 standard; DNA; 3771 BP.
XX
AC Q51689;
XX
DT 09-DEC-1995 (first entry)
XX
DE Bacillus thuringiensis crystal protein 33f2 gene.
XX
KW Crystal protein; delta endotoxin; acaricide; pesticide;
KW biological control agent; transgenic plant; crop improvement; ds.
XX
OS Bacillus thuringiensis (isolate 33f2).
XX
FH Key Location/Qualifiers
FT mat_peptide 1..3771
FT tag a
FT misc_feature 4..24
FT tag b
FT function= oligonucleotide hybridization probe
FT /note= "Probe A"
FT misc_feature 13..33
FT tag c
FT function= oligonucleotide hybridization probe
FT /note= "Probe B"
XX
US5262158-A.
XX
PD 16-NOV-1993.
XX
PF 30-APR-1991; 91US-0693210.
XX
PR 30-APR-1992; 92US-0876280.
XX
PR 30-APR-1991; 91US-0693210.
XX
PR 13-SEP-1991; 91US-0759248.
XX
PR 30-SEP-1991; 91US-0768141.
XX
PA (MYCO ) MYCOGEN CORP.
XX
PI Bagley AL, Cannon RJC, Payne JM;
XX
DR WPI; 1993-377387/47.
XX
DR P-PSDB; R44203.
XX
PT Controlling acaride pests e.g. two spotted spider mite - utilising
XX Bacillus thuringiensis delta-toxins cloned into microbe hosts.
XX
PS Disclosure; Column 43-48; 42pp; English.
XX
CC DNA encoding the insecticidal toxin can be cloned into baculo
CC viruses and transferred to other host microbes, preferably E. coli
CC NM522(pMYC 2316) NRRL B-18785, to control acaride pests, or to
CC plants which become resistant to the acaride pests.
CC Specifically, the two-spotted spider mite (Tetranychus urticae) is
CC controlled.
XX
SQ Sequence 3771 BP; 1417 A; 614 C; 591 G; 1149 T; 0 other;

```

alignment\_scores:

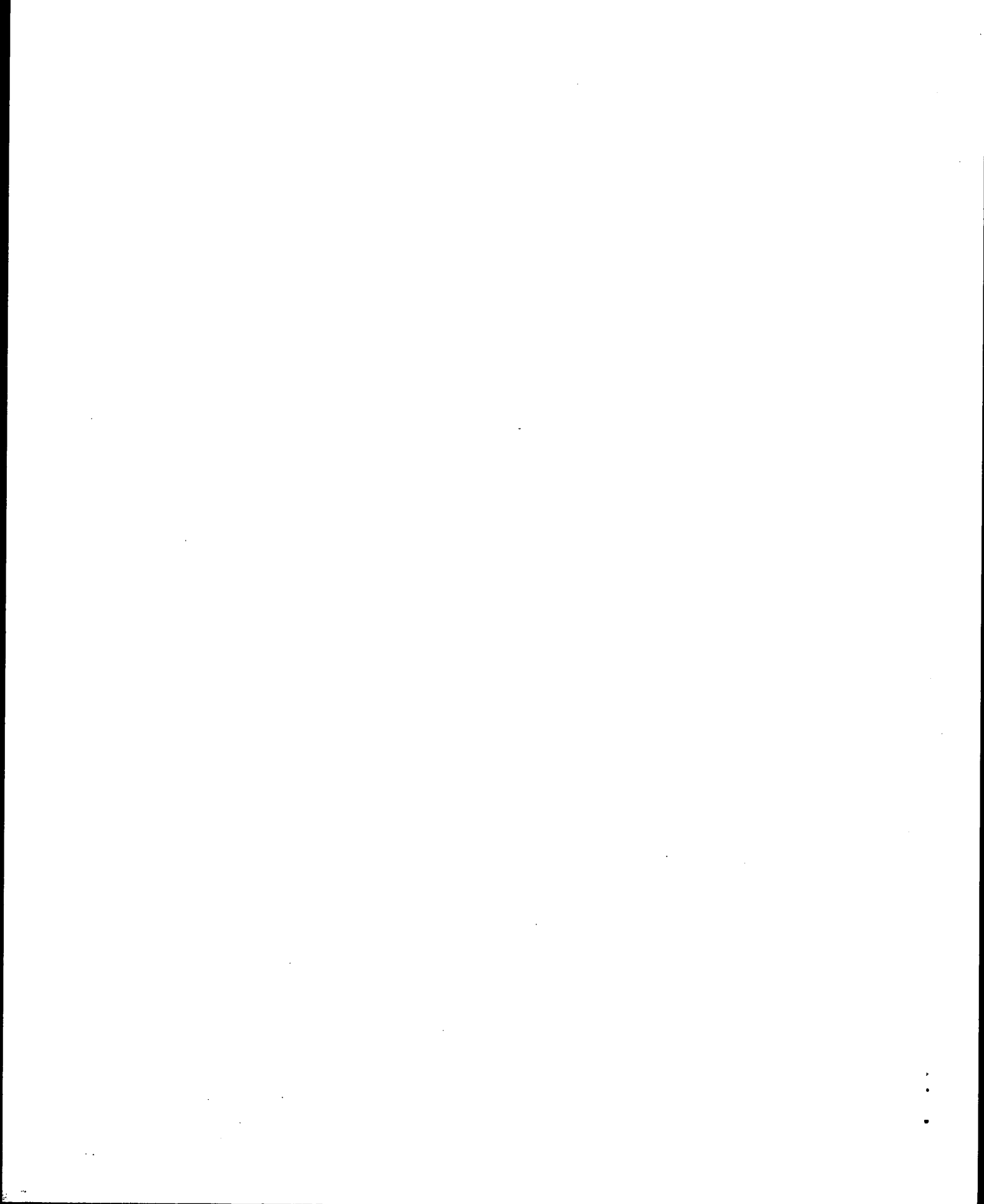
```

Quality: 107.00 Length: 491
Ratio: 0.493 Gaps: 22
Percent Similarity: 44.196 Percent Identity: 19.552
alignment_block:
US-09-426-072-2 x Q51689 ..
Align seg 1/1 to: Q51689 from: 1 to: 3771

6 LeuLeuThrGlyLeuGlyLeuAlaSerLeuGly...LeuAlaSerPr 21
: : : : : : : : : : : : : : : : : : : : : : : : : : :
220 GTTTTATCTGCTAATTAACCTTGAAGTTGCTACTGTGCTCTGTC 269
21 oValThrSerGluTyrThrSerValArgGluAlaProPheGlyTyrLysP 38
: : : : : : : : : : : : : : : : : : : : : : : : : : :
270 AGCAAGTACTATTGTAAGTTTATTGGCTTAAATATTGGAGATAAAC 319
38 roGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIle 54
: : : : : : : : : : : : : : : : : : : : : : : : : : :
320 CAATGCAAAAATATATTTGAAGAGCTCAAGGCTCAAAATTTGAACATTA 369
55 ValTyrLeuLeuLeuGluAsn.....ArgSerPh 64
: : : : : : : : : : : : : : : : : : : : : : : : : : :
370 ATTCACAGATATATAACAACTATCAAGATGCAATTAATCAAAAAAATT 419
64 eaAspAsnIleLeuGlyValValArgGlnGlyLeuAspAsnProIleA 81
: : : : : : : : : : : : : : : : : : : : : : : : : : :
420 TGACAGCTTTCAGAAAACANTTAATCTATATACAGTA.....GCTATAG 463
81 snAsnGlyProPheCysAsnTyrLysAsnAlaSerAspProSerSerGly 97
: : : : : : : : : : : : : : : : : : : : : : : : : : :
464 ATACAATGATTAACGTAACACAAAA.....489
98 LysTyrCysThrGlnAlaLysAspTyrAspSerValPheAsnAsp..... 112
: : : : : : : : : : : : : : : : : : : : : : : : : : :
490 .....ACGCAACTCGAAATCTAAATCTATATCTTACCTACCATAT 530
113 .....ProaspHisSerValThrGlyAsnAsnLeuGluPheT 125
: : : : : : : : : : : : : : : : : : : : : : : : : : :
531 CTCCATATTTATCCAGAGGATATGAAACTGGA...GGTTTACCTATT 577
125 yrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValVal 141
: : : : : : : : : : : : : : : : : : : : : : : : : : :
578 ATCTTAATGTTGCTAATGCTCATATATATTGTTAAGAGACGCTATAGT 627
142 ...AlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLy 157
: : : : : : : : : : : : : : : : : : : : : : : : : : :
628 AATGCAGAGAAATTAGGCTTTAGTGATAAAGAGACTAGACACACATAAAA 677
157 stLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGluG 174
: : : : : : : : : : : : : : : : : : : : : : : : : : :
678 ATAT .....ATCAAAATGCACANTACACATCATACTGAAG 712
174 luGluValProThrLeuValAspLeuValAspGluPhe.....186
: : : : : : : : : : : : : : : : : : : : : : : : : : :
713 CAGTAATAAAGCATTTCTTAATGGACTTGACAAATTTAAGAGTTTAGAT 762
187 ...ThrThrPheAsnSerTyrPheSerCysValProGlyProThrAsn.. 201
: : : : : : : : : : : : : : : : : : : : : : : : : : :
763 CTAATAGCTATATAAATAAGCAATATATTAAGGTATGACAGAAAT 812
202 .ProAsnArgLeuCysAlaLeuAlaGlyThr...AlaAlaGlyHisGlyL 217
: : : : : : : : : : : : : : : : : : : : : : : : : : :
813 GGTCTTGATCTAGTTGCTCTATGGCCAACCTTCGATCCAGATCATATC 862
217 ysAsnAspAspPheLeuAsnTyrGlyLeSerSerLysSerIlePhe 233
: : : : : : : : : : : : : : : : : : : : : : : : : : :
863 AAAAAGAGTAGAATTAATTTTCAAGAACTATTCTCTCCAATTAC 912
234 GluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAs 250
: : : : : : : : : : : : : : : : : : : : : : : : : : :
913 CAACCTGTACCTAAA.....AACATGCAAAATACCTC 944

```

250 nGlyuPheGlUpProAsnSerLeuPhe..... 259  
: : : : : ||| : : : : :  
945 TAGCTCTATTGTACCTAGCGATCTATTTCACCTATCAAGGAGATCTTGTA 994  
:  
260 .....Phe 260  
|||  
995 AATTAGAATTTTCTACAGNCGCACAGCATGGTCTTCGCAAAATTTTT 1044  
:  
261 ThrTyYValAnGlnThrSerArgSerAsnValProValGluAsnPh 277  
||| : : : : : ||| : : : : :  
1045 ACTGGTATTCCGAACACATCTTACAACATCGCCTAATACTCATCAACACA 1094  
:  
277 ePheGlnAspAlatyrLeuGlyValLeuProLysPheSerTyYrIleAsn 294  
||| : : : : :  
1095 CCATGTGATAT.....TTTAGTTAT..... 1113  
:  
294 roSerCyScySGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 310  
||| : : : : : ||| : : : : :  
1114 .....AATACCCAATCT.....AGTGGTAATAT 1137  
:  
311 SerTyYrGly..... 313  
||| |||  
1138 TCAAGAGGCTCTTCAAATCCGATTCCAATGATCTTAATAATCCCATTAT 1187  
:  
314 .....GluValPheVallysGlnIleTyRasp..... 322  
: : : ||| ||| |||  
1188 TTCACCTGTATTAGAAATTCATTTTATAGGCATACGCCGATCTTCG 1237  
:  
323 .....AlaIleArgGlnGly 327  
||| : : : : :  
1238 TTTTGTAGTTAATTTTAAAGATGSCACTCAAGGATATGCATTTGCCCAAGCA 1287  
:  
328 ProGlnTrpAspLysThrLeuLeuPheIleThrTyRaspGlnThrGlyCl 344  
||| ||| ||| ||| ||| ||| |||  
1288 CCA.....ACAGGAGG 1298  
:  
344 yPheTyRAspHis.....ValProProLeuAlaValArgProAspA 359  
| : : : : : ||| : : : : :  
1299 TGCGTGGGACCATTCTTTTATTTGAATCTGATGGTCCCCAGAAGGCGATA 1348  
:  
359 snLeuThrTyRThrGluThrAlaLysasnGly..... 369  
: : : : : ||| : : : : :  
1349 ANTTAAACTATATTATATCTCTCCAGGTGATACATTTAGAGATTTTCATC 1398  
:  
370 GlnLysTyRThrLeuHisPheAspArgLeuGlyArgMetProThrTr 386  
: : : ||| ||| : : : : :  
1399 ANTGTATATATCTTATATAGTACTCCAACCTATAAATGAACATATCAACAGA 1448  
:  
386 pValIleSerProtyr...SerLysLysGlyTyRilecluInlTyR 402  
||| : : : : : ||| : : : : :  
1449 AAAATCAAAAGCTTTCCTGCGGAAAAAGGATATATCAAAAATCAAGGGA 1498  
:  
402 hrAspProValThrGlyLysPro 409  
||| ||| ||| ||| ||| ||| |||  
1499 TCATGAATATATTACGGTAATCAAC 1521





```

/protin_id="AAF82193.1"
/db_xref="GI:8954019"
/translation="MGVAVATSTTHQWVAHPIVHSSSLSSQTLASSAVSSRRRII
GNDGRSLSCRSVQMSQSLSPFSTNLHRSKSCSELWETKRPKQLIRRAFA
NYDFSEEEFAKKMOELTLKQVSNEDNESDTRIDYTRKMDIVGSHNNFRSDME
PWPMVOMSNIERKANSVLPPLSLRIKRLKLOMBEGLVNGVGSACCSKAPSSVA
FMIRELOSTLHMRLELFDLOGLIHRVKEQWASFWVLLFQVFSATPTLAVSVMLL
LANFTVYSIESNSALAAVSPPTLSFSFTI"
5532..9581
/gene="F10K1.2"
/notes="F10K1.2"
Join(6532..6753,6846..6939,7020..7240,7328..7402,
7474..7588,7861..7920,8036..8109,8235..8285,8402..8451,
8640..8721,8824..8913,9041..9103,9225..9320,9453..9581)
/gene="F10K1.2"
/notes="Contains similarity to a putative Cdc6 protein
At2g29680 g|3582344 from Arabidopsis thaliana BAC T27A16
gb|AC005496."
/codon_start=1
/evidence-not_experimental
/protein_id="AAF82194.1"
/db_xref="GI:8954020"
/translation="MPTNAGTSLSSYKHIVAIGTTVKSESLESAAYEIPRKRMSDS
AAVGSNSVTPKKLKHLPSPVPMGMSKVEVDSNEILRPVNLAVSDCLGTSKW
SPREOMRAVKEALHVSAPSTLRCDEQIRIFEFVKGCDIQKAGSLYICGPGT
KSLSKKVVQVQWSTQAGLPVDTLSVNTSLSKTIDIFSKLIGELKPKGNANX
SSPHQLQLRFSQKQSSSSMMYGVANAIDLAADRFPLKLSLCKPMWITFRAYSD
QILRLOERLRLSVYAFQKALELCAKVAASGDMKALCVCRSALEILEITRGS
TGPESGQTPDDSVVMDHMAALSKTFKSPVETIQSLPHQOIIICAAAKAFRGSK
KDATVGLKLYLEICKSMWISFAGITEFTNCTVLDNDQGLKVGQARRDKLRVSLR
VDESDFALQVNSLILFPAN"
10251..11681
/gene="F10K1.3"
/notes="F10K1.3"
Join(10251..11681)
/gene="F10K1.3"
/notes="Strong similarity to an unknown flavonol
3-O-glucosyltransferase At2g29740 g|3582341 from
Arabidopsis thaliana BAC T27A16 gb|AC005496. It contains
a UDP-glucuronosyl and UDP-glucosyl transferases domain
PF100201. ESTs gb|746737, gb|A1993247, gb|T76043,
gb|AV550669, gb|AV538399 and gb|AA720097 come from this
gene."
/codon_start=1
/evidence-not_experimental
/protein_id="AAF82195.1"
/db_xref="GI:8954021"
/translation="MKAEEIIFVTPSPGHLVSIIEFAKSLIKRDRDRIHTITLYVA
LPLAQHLFAKSLVASOPRIALLPDVQNPPLLEFFRAPEAYILESTKTKTVLPR
DALSTVSSRKESGSRVVGVLVDFPCVPMLEVANLNPSTFTFCNAGLSMMKYI
PERHRTTSELDSLSSGNVEHPIDPVYCVSPVKVLPGLFVRESYFAWVEIAEKFPGAK
GILVNSVTCLEONAFDARLDENPPVYPCVPLSLKDRSPNLDASDRDRIRWLE
DOPESSIVICFSGSLIIGKLOIEETAEALTLHRLFMSIRINPTERKASPTDLLPG
FLURTASGLVCDNAPQVEVLAHKAUGFVSHCGWNSVLESIFWFGVPIATWPMYAEQQ
LNAPSMKELGLAVELRLDYVSAYGEIVKAEIAGAIRSLMDGEDTPRKRKVENAAE
RNALMDGSSSFVAVKRFDELIGGV"
complement(12027..13466)
/gene="F10K1.4"
complement(12027..13466)
/gene="F10K1.4"
/notes="Contains similarity to an unknown flavonol
3-O-glucosyltransferase At2g29750 g|3582329 from
Arabidopsis thaliana BAC T27A16 gb|AC005496. It contains
a UDP-glucuronosyl and UDP-glucosyl transferases domain
PF100201. ESTs gb|A1993795, gb|N97301 and gb|Z18063 come
from this gene."
/codon_start=1
/evidence-not_experimental
/protein_id="AAG18592.1"
/db_xref="GI:10567858"
/translation="MKVETELIPVPSTGHILVHIEFAKRLINLDRHRIHTITLNL
SPSSHASVFARSLIAQKRIHLDPIDQPPFDLYQRAPEAYIVKILKNTPLIK
YIPDRHRTASEFDLSSGDELVPQFVFNAPTKFMPGGLFNKAEYAYELAPLGMK
AKGILNVSFTELEPHEDYFSLHEKFPVPPVPGPILSKDRASPNEEAVDRDQVGLW

```

```

gene
CDS

alignment_scores:
  Quality: 435.50
  Ratio: 1.853
  Percent Similarity: 61.358
  Gaps: 14
  Length: 383
  Percent Identity: 32.115

alignment_block:
  US-09-426-072-2 x F10K1
  Align seg 1/1 to: F10K1 from: 1 to: 104533

```

```

DDPESSVVFCLGSGRSVDPOVKETARLELVGCRFLMSIETSGDVTNPNDVLP
GFMRVAGRLGCGWAPQVEVLAHKAUGFVSHCGWNSVLESIFWFGVPIATWPMYAEQ
QLNATLVKELGLAVDLMDYVSRGLVTCDEIARAVRSLMDGGDEKRRKRVKEMADA
ARKALMDGSSSLATARFIELFEDQSSC"
complement(14098..15540)
/gene="F10K1.5"
complement(14098..15540)
/gene="F10K1.5"
/notes="Contains similarity to an unknown flavonol
3-O-glucosyltransferase At2g29750 g|3582329 from
Arabidopsis thaliana BAC T27A16 gb|AC005496. It contains
a UDP-glucuronosyl and UDP-glucosyl transferases domain
PF100201. ESTs gb|A1997635, gb|T13644, gb|AV546216 and
gb|AV596826 come from this gene."
/codon_start=1
/evidence-not_experimental
/protein_id="AAG18591.1"
/db_xref="GI:10567857"
/translation="MKTAEILFVPLPETGHLILTETFGKLLNLDRIITLITLML
PYAHADASLASLTASPGIRIILSLPEIHDPPIKLLDTSSEYIILDFIHNKPLCKR
TIQDLVSSSSSGGSHVAGLIDDFCVGLIDIGREVNLPSTYIFMTSNFGLVLOV
KILRLNLTSEEDSGEELHPIAEVNRVPAKVLPPGVFKLSGLSVKIGEMKHEA
KGLIVNSFTVEPYAAEHFSQGRDYPHYVPVNLNLTGRNPLGSLAQYKEMMKWLD
PODSSVLELCFGSMGVFPAPQITEIAHALEIGCFWIAIRTNMAGDQDPOEPLPG
FVDRTMGRGIVCGWAPQVILAHKATGCGFVSHCGWNSVLESIFWFGVPIATWPMYAEQ
LNAPSMKELGLAVELRLDYVADGDRVTLVISADEIATATVRSLSMDSDNPVRKVKIEK
SSVARKAVGDKGSSVTATCNCFKIDILGDHF"
16651..18920
/gene="F10K1.6"
Join(16651..17749,18047..18223,18595..18920)
/gene="F10K1.6"
/notes="Contains similarity to an unknown protein T0P19.120
g|16523092 from Arabidopsis thaliana BAC T0P19
gb|AL133315. ESTs gb|84021, gb|A1992399, gb|H76814,
gb|F15169, gb|AA585873, and gb|AA605516 come from this
gene."
/codon_start=1
/evidence-not_experimental
/protein_id="AAF82197.1"
/db_xref="GI:8954023"
/translation="MAFRVLTIVLCYLISSQSEFKNOKPHKIQGPIKTVV
VMENSFHILGWLKSTRPEIDGLTGKESNPLVSDPNKIFVSDAVFVMDPGHS
FOAIRQIFGNDTSGDKPMNGFAQOSSEMPGMKKNVSGKPEVLPVYTELANERG
VDFRNFASVPTSTQPNRFVHSATSHGCSNVKDLVKGPDKTDFDSLDEGLSGCI
YONTPAFVTEFKSLRLHLVKFHSALKKDLAKLGN*SVYEQRFDDIDLFPAN
DDPSHDAAGORVKEVETELSSPOWKEMALLITDEGGFYDHVPTPKVGNPD
GIIGDPDVFDFGLRVETELISSKIEKTVIHEPGETPHSQFHESSIPATVKKL
FNLSHFLTKADAWAGTFEYFRIDSRQDCEPKLPEVKLSRPMWGADESKLSEFO
VELIQLASOLVGDHLNLSYDPIDGKNMTVSEGNKYAEDAVQKLEAGMAALEAGADENT
IVTMRSPLTFTSPSTGNTKYIGSY"
20050..22356
/gene="F10K1.7"
Join(20050..20169,20522..20760,20877..21227,21304..21493,
21570..21634,21720..21930,22009..22356)
/gene="F10K1.7"
/notes="Contains similarity to an unknown protein T2J13.180
g|16522368 from Arabidopsis thaliana BAC T2J13
gb|AL132967. ESTs gb|Z29835 and gb|Z29836 come from this
gene."
/codon_start=1
/evidence-not_experimental
/protein_id="AAF82198.1"

```



```

34  PheGlyTyrLysProGlySerLysGluSerIleGlu..... 45
16687 TTCTGCTACTACTAATTTTCAATCCATTCATTAAGTTTCAAAACATCCCA 16736
46  ....AsnLeuLysAspLysValGluAsnIleValTrpLeuIleL 59
16737 GAAACCCCAAGATCCCAAGGACCAATCAAGACTATAGTCGTTGTAGTCA 16786
59  euGluAsnArgSerPheAspAsnIleLeuGlyValArg.....Arg 73
16787 TCGAATCGCTTCCTTCATCATCTCGGTGGCTCAAAATCGACCCGA 16836
74  GlnGlyLeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrLysAs 90
16837 CCCGAATCGACGA.....TTAACCGGCAAGAGTCAATCCCTCAA 16880
90  nAlaSerAspProSerSerGlyLysTyrCysThrGlnAlaLysPtyrA 107
16881 CGTCTCGATCCAAATTCCTCAAAAGATCTTCGTCACAC.....G 16921
107 spSerValPhe.....AsnAspProAspHisSerValThrGlyAsnAsn 121
16922 ACGCGCTCTTCGTAGCATGATCCGGTCACTCATTCCTCAAGCGATCCGA 16971
122 LeuGluPheTyrGlyThrTyrProAsnAsnGlyAlaIleAlaSerG 138
16972 GAACAGATATTCGG.....TCTAACGATACATCCGGCGACCC 17009
138 yLysValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAsp 155
17010 GAAATG.....AACGATTCGACAGCAATACAGAGATATGG 17047
155 yrProLysLeuAlaProGluGluAlaThrArgGlnValMetGlyTyr 171
17048 AGCTCGAATGCT.....AAGACGTTATGACGGGCTTC 17082
172 ThrGluGluValProThrLeuValAspLeuValAspGluPheThr 188
17083 AAACCCGAGTCTTACCGGTTTACCGAGTTAGCGAACGAGTTCGGAGT 17132
188 rPheAsnSerTyrPheSerCysValProGlyProThrAsnProAsnArg 205
17133 TTTCGATCGATGTTTCGTCAGTACCTACCTCGACTCAACCAATCGGT 17182
205 euCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsnAspAsp 221
17183 TCTACGTTCACTCACTACTTCTCATGGATGCTCAAGCAACGTGAAGAA 17232
222 PheLeuAsnTyrGlyIleSerLysSerIlePheGluAlaAlaAsnG 238
17233 GATCTCGTTAAAGGTTTCTCAAGAGAGATCTTGATTCGTAGTCA 17282
238 uLysGlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyCluPheG 255
17283 GAACGAGTCACTTT.....GGGATTTATPACCAAAACATCC 17320
255 roAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSerAsnVal 271
17321 CTGCGAGTCTTCTCTCAAGTCTCTTCGTAGGCTT.....AAGCATTTG 17364
272 ValProValGluAsn.....PhePheGlnAspAlaTyrLeuG 284
17365 GTGAATTCATAGTATGCTCAATGATTCAGCTCGACGCTAAGCTTG 17414
284 yValLeuProLysPheSerTyrIleAsnProSerCysGlyThrAsn. 300
17415 GAAGCTCCGAATATTCTGTCGTGGACAGAGATATTCGATATCGATC 17464
301 .....ThrAsnSerMetHisProThrGlyAsnValSerTyrGlu 314
17465 TTTTTCGCGCTAATGATGATCATCCGTCGATGCGCGCGGCTCAA 17514
315 ValPheValLysGlnIleTyrAspAlaIleArgGlnGlyProGlnTrpAs 331

```

```

17515 CGGTTTGTAAAGAGTTTATGACAGCTTACGAGTAGTCCACAGTGAA 17564
331 pLysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheTyrAsp 348
17565 AGATGATGGCTTGTCTTACTTATGATGAGCAGCGTGGTTTATGACC 17614
348 isValProProLeu.....AlaValArgProAspAsnLeuThrTyr 362
17615 ATGTTCCTCCACGCGGTTAAAGGTGTCCCTAACCCCTGATGGTATCAT 17661
363 ThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAspArgLe 379
17662 .....GGACCTGATCGCTTTTACTTTGGTTTGACCGATT 17696
379 uGlyGlyArgMetProThrTrpValIleSerProTyrSerLysLysGly 395
17697 GGGTGTTCGGGTCCTACCTTCTGATCTCTCTCGATTGAGAAGGC 17745

seq_name: gb_p12:ATT8P19
seq_documentation_block:
LOCUS   ATT8P19          96679 bp    DNA             PLN      02-DEC-1999
DEFINITION   Arabidopsis thaliana DNA chromosome 3, BAC clone T8P19.
ACCESSION   AL133315
VERSION     AL133315.1  GI:6523080
KEYWORDS    thale cress.
SOURCE      Arabidopsis thaliana
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons, core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis
REFERENCE   1 (bases 1 to 96679)
            Choinsne, N., Robert, C., Brottier, P., Wincker, P., Cattolico, L.,
            Mayer, K.F.X., Saurin, W., Weissenbach, J., Mewes, H.W., Lemcke, K.,
            Artiguenave, F., Quetier, F. and Salanoubat, M.
            Unpublished
JOURNAL    Direct Submission
REFERENCE   2 (bases 1 to 96679)
            EU Arabidopsis sequencing project.
AUTHORS    Submitted (02-DEC-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemccke@mps.blochem.mpg.de, mayere@mps.blochem.mpg.de Project
            Coordinator: Marcel Salanoubat and Francis Quetier, Grouperment
            d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France;
            http://www.genoscope.cns.fr
COMMENT    Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.blochem.mpg.de/proj/thal/.

FEATURES             Location/Qualifiers
     source            1..96679
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
     gene              36..963
                     /chromosome="3"
     exon              36..86
                     /gene="rP19.10"
                     /number=1
     CDS               join(<36..86,324..475,546..963)
                     /gene="rP19.10"
                     /codon_start=1
                     /product="hypothetical protein (fragment)"
                     /protein_id="CAB62339.1"
                     /db_xref="GI:6523081"
     translation       "KTSVKELDEAALTEROHNKLMWEAKARGEGYKIDLRNIDM
            DEYDFLHWRRLSLEALLDISSQALGLPLEPGFKGFGDYDPTSALOY
            DYWGEPKSEISKQDMKDAHNSIVGKNWYDMSYDDAIKOTIEKRKGSTLASQE
            EETSEEEEDDDDFDDFDYSILSDESSIGYSEQQLVNGTQVLTD"
     intron            87..323

```

exon	/gene="T8P19.10" /number=1 324..475	exon	23114..23318 /gene="T8P19.40" /number=1
intron	/gene="T8P19.10" /number=2 476..545	intron	23319..23700 /gene="T8P19.40" /number=1
exon	/gene="T8P19.10" /number=2 546..963	exon	23701..23873 /gene="T8P19.40" /number=2
exon	/gene="T8P19.10" /number=3 964..10231	intron	23874..23991 /gene="T8P19.40" /number=2
gene	complement(3123..4007) 3123..4007	exon	23992..24147 /gene="T8P19.40" /number=3
CDS	complement(3123..4007) /codon_start=1 /product="hypothetical protein" /protein_id="CAB62340.1" /db_xref="GI:6523082"	intron	24148..24240 /gene="T8P19.40" /number=3
trna	/translation="MDGRGCCIAARYAIGSPYDLSKADRTMLRERPIAPKSPGGV NPVSSGSGSSDVSRFSASRRKCHOLKENGNAKRCRPTKSDKPVHGSDNAV TLLSLPKPDLKAAVEKQKQGLPLWSKDGGMLETAYQTPEIVQRTVIVISSCMTV ERVDAWIDYGLGRSDEERMLNVRDTCPEISDGRVTVNDAYRKMAADIIPE EGAPEITSGDSFHVIVLRMRPMLTSPGTCRMKLOYTCQNRGRSVTVPCDWRM DVGGFAWRLDVKAAALC"	exon	24241..24516 /gene="T8P19.40" /number=4
gene	/note="Tyr-tRNA: anticodon GTA" 10659..12179	intron	24517..24594 /gene="T8P19.40" /number=4
exon	/gene="T8P19.30" complement(10659..12179)	exon	24595..24758 /gene="T8P19.40" /number=5
CDS	complement(10659..12179) /note="strong similarity to cytochrome P450 CYP94A1 - Vicia sativa, PIR2:T08014; contains EST gb:H76099" /codon_start=1 /product="cytochrome P450-like protein" /protein_id="CAB62341.1" /db_xref="GI:6523083"	intron	24759..24846 /gene="T8P19.40" /number=5
misc_feature	/translation="MAFLLSLIALLIITIFLSSSTKKVOENTYGPSPYPLIGS ILSPNKNRHLQWTELLRISPSOTILPLIGNKRIITINPLNVEIILKTNFNFP KGFPTDLGLDGGIFNVGHSQSKLASHEFSTRSLRSPAFELVLDKVENRV PVLSTADGTTVDQVLKRFADVVCVKSGLWDPCDLTRFVPLVEAFDTAAEI SARATEPIYAVNKTRVLYNGSERKLEAIRTIVHLYSEIVRAKKSLKLEIGTGAEK QDLISREFLAHNGEAVRDWISFIMAGRDYTTSAAMTWLWLLTENDDVERKILEVD PLVSLGLGFEDLKEMAYTACILCEAMRLYPVPSWDSKHAANDVLPDGTVRKRGVKT YFPYGMGRMETLWGTDEENPNKWFDESPGSTRPVLPKIPSPYKFPVFGAGPRVCVK EMAFOMKYVVGVSLSRFEIIVPNKDRPVFVPELLTAHMAAGLKVYIKRSHILNV"	exon	24847..25147 /gene="T8P19.40" /number=6
gene	14300..14811	exon	complement(25511..25630) /number=1
misc_feature	14300..19503	gene	25511..27570 /gene="T8P19.50" complement(join(25511..25630,25830..26124,26534..26592, 26686..26775,27173..27248,27380..27570))
gene	23114..25147	CDS	/note="EST gb:234027 marks introns at position -27177, 26775-26736, 26592-26534, 26417-26380, 26213- without corresponding gene model; similarity to deoxycytidylate deaminase - Homo sapiens, PIR:I55434; contains EST gb:234027" /codon_start=1 /product="putative protein" /protein_id="CAB62343.1" /db_xref="GI:6523085"
CDS	join(23114..23318,23701..23873,23992..24147,24241..24516, 24595..24758,24847..25147) /gene="T8P19.40"	intron	/translation="MNSRDLTVSTSAIFGALISALAFRFFSSNPNKPKRFTSTET SAISRFPALDPFSLKRNGLYSDDDYFMATFALSAERSKDPNRQVAGCLVSQNGVIL GKTIIRSLHCFDLFEQHEGLAIMVFOEALMTNPLGSPYFLTNHGVSGLSERIKTFKV SYTRISSEYSLIGFCANKLEMLFGIENLNSAAVSPKYLCDGLYEFVICSNNLL QKLYVTMFPNCAKILIQSGVAEVIYFVEKRPNDSVDVAVVASHKLLSMANVVKIKAT "
misc_feature	/note="EST gb:742970 spans first intron, EST gb:N96487 spans third intron; similarity to probable transcription regulator protein - Sorghum bicolor, PIR2:T03446; Contains Prokaryotic membrane lipoprotein lipid attachment site AA347-357"	exon	complement(25631..25829) /number=1
gene	/codon_start=1 /product="putative protein" /protein_id="CAB62342.1" /db_xref="GI:6523084"	exon	complement(25830..26124) /number=2
CDS	/translation="NATVPEIKIMRSESLGHSRSDVSSPEAKLGMRVEDLWDEQKPLS FNEKLACFESIPVSAFPLSSDSQDIERSDTSLAEAQVLSKFKVLSAPVVDVDAPE DASMDRYIGIVEFFGIIVWLLHQLPEPPSPRPAVAANSFSDHFTDVLVDNGAVT SGNFVEITYSSELYKNTKRDISCTFRWAPFLAQKENSFLTMILLLSIKMKKSIPTV DLGVAKENITQSGVIHMLAEACGLWFEDWGIKTLSEVLGPIMSKDHIIKIYDEDP VLOAQFKLMRKRIIGIPVIERNSEKPVGNISLRDQVFLITAPEIYHYDSITTKNFI SVRHELEKCGDTSAPIMSGVIACITKNITKELILMLDAEKIHTRIYVVDVDFGNLEGLIT LRDIIARLVHEPFGYFGDFDGVNPLPNRY"	intron	complement(26125..26533) /number=2
		exon	complement(26534..26592) /number=3
		intron	complement(26593..26685) /number=3
		exon	complement(26686..26775) /number=4
		intron	complement(26776..27172) /number=4
		exon	complement(27173..27248) /number=5
		intron	complement(27249..27379) /number=5
		exon	complement(27380..27570)



Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr.  
Rockville, MD 20850, USA  
email: xlin@tigr.org  
BAC clone T21P5 is from Arabidopsis chromosome III and is near the  
molecular marker ml172.  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.

Genes were identified by a combination of three methods: Gene  
prediction programs including GRAIL (available by anonymous ftp  
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of  
Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/~chris/genscan.html>), and NetPlantGene  
(<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the  
complete sequence against a peptide database and the Arabidopsis  
EST database at TIGR (<http://www.tigr.org/tdb/at.html>).  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as 'unknown' proteins.  
Genes without protein or EST similarity, that are predicted by more  
than two gene prediction programs over most of their length are  
annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
predicted by tRNAscan-SE (Sean Eddy,  
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are  
identified by repeatmasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of  
genomic sequence that are not annotated as genes but have predicted  
exons by GRAIL are annotated as misc features.

#### FEATURES

```

source
1..82289
  /organism="Arabidopsis thaliana"
  /cultivar="Columbia"
  /db_xref="taxon:3702"
  /chromosome="III"
  /map="ml172"
  /clone="T21P5"
misc_feature
1..24217
  /note="Overlap with BAC clone T12J13
  (AC009327:65718..89934)."
  complement(join(<8..215,451..587,675..857,959..1050,
  1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
  2979..3100))
  /gene="T21P5.1"
  complement(<8..3100)
  /gene="T21P5.1"
  /note="similar to hypothetical protein GB:CAB38918
  [Arabidopsis thaliana]"
  complement(join(8..215,451..587,675..857,959..1050,
  1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
  2979..3100))
  /gene="T21P5.1"
  /note="unknown protein"
  /codon_start=1
  /protein_id="AAF01600.1"
  /db_xref="GI:6017117"
  /translation="MSQIISGNNNIPSEVYVSLVKNADKKFKIRDLFPYERGRYEN
  YFFKFKVYVTLQWQENRQKLEAGLKEIGETIASRIAQLYYGYHYRTSDAGYLS
  ESYVFEAILITREYEDGLFODLIANKQLREFLRLMVLGRLREVMYHOLVQFKR
  LIDCKRTQETDEKFEKVAQELVRELFKSDTAFPMNIRPLRSLVLDPLNDLQTPRAS
  RSLRLTDAILSSYCNVYKSELTLDLDFRMLOCLEWPEPSGLYQSGAKGNAPGVG
  ARINSQSDMLPLPNPKAVLYRPSITFLAVLATICEELPSHGILLILYLSAGSKIG
  QISSPLSARSATSVENILLDRESWTIKQETESLQITPSGSSYIYPSDLVPFTFRAP
  LFIILSDSTVFNKICGAEGKGEPAALLSPHPLPLISADFSRQPSGSLFIETLSP
  VQAFLLISASDMDIETFKAEKLLSSMNEWASTLSDTLHPVNSQILKQFFLR
  RLLRFIFRCVRLALYTPVENNKKQPECCPSLPSLPTAPAVQSAFQNAVNFAT
  SKFTIPQDITMLESF"
  join(<3627..3661,3762..3843,3918..3995,4081..4140,
  4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
  5361..5494,5576..5682,5785..5827,5903..6144)
  /gene="T21P5.2"
  <3627..6144
gene
```

```

/feature="T21P5.2"
  /note="predicted by genscan"
  join(3627..3661,3762..3843,3918..3995,4081..4140,
  4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
  5361..5494,5576..5682,5785..5827,5903..6144)
  /gene="T21P5.2"
  /note="hypothetical protein"
  /codon_start=1
  /protein_id="AAF01580.1"
  /db_xref="GI:6017077"
  /translation="MDDNRSSSKRHEIEKDTIASRKLEDNTKLIQDPERMALYAK
  VRSEIEHSIERAAACQLMQLNERYGLERKADLRVAIDEKQNSVTSALNEL
  ARRGDLEENKLAHLDEYEDRYIFMTSLGLLAEGYVMPRVANATAISSGKHLH
  DQLOKTKACNDRIRRELSIVENQPGTDFISKNDHDPNSKTSOASGSDRNDYQTN
  BQLLPMPNVTNPNYINIMODTESLRNNOIGGSGQIFPOPKRENGVPLSSVAGKE
  MIQREKAEANSMPFDAYNGNEPASHVVEEGPGIDGFOIIGDAIPCERKVLGGCPVR
  GTTLCMFQVRLHEDGTROYLEGATPEIYVADDDVKLIAVECIPIPDQGRQVAYRD
  ESGIYSENSVSKVDLLIMQGLVRLFANDQNKIRCGNVNGT"
  6225..6289
  /note="exon predicted by xgrail, quality marginal"
  6384..6503
  /note="exon predicted by xgrail, quality excellent"
  6794..6839
  /note="exon predicted by xgrail, quality excellent"
  6928..7033
  /note="exon predicted by xgrail, quality marginal"
  7293..7411
  /rpt_family="(TAAAA)n"
  complement(7488..7547)
  /note="exon predicted by xgrail, quality
  excellent shadowexon"
  complement(7654..7784)
  /rpt_family="(TAA)n"
  complement(7819..7864)
  /note="exon predicted by xgrail, quality good shadowexon"
  8054..8113
  /note="exon predicted by xgrail, quality marginal"
  8263..8380
  /rpt_family="(TAAAA)n"
  complement(8592..8627)
  /note="exon predicted by xgrail, quality
  marginal shadowexon"
  <8652..>9527
  /gene="T21P5.3"
  <8652..>9527
  /gene="T21P5.3"
  /note="contains zinc finger motif, C3HC4 type (RING
  finger)"
  8652..9527
  /gene="T21P5.3"
  /note="unknown protein"
  /codon_start=1
  /protein_id="AAF01602.1"
  /db_xref="GI:6017119"
  /translation="MGSTGNPNWGTYYDSVDCSQGVCSVYCPWCYVIFPPPPSY
  LDDDDSSDFSPLLIAGILASAFILVSYTLISKYCHRRHNSSTSAANIRI
  SSYTWQGTNNNNNGATNPNOTIGGGGDLGSLIKITVYVKRMDFVSSDCS
  VCLSEFQENSLRLPKCNHAFHPCIDTWLKHSCNCPICRAFTVTSAAVETDLTQ
  QIVTNNISITGDDSVVVVNLDSNRSRNETVNEGSTPTXGADGPKRRRAEIGVVK
  LRRCGIDRRYTKGD"
  complement(10898..10896)
  /note="exon predicted by xgrail, quality
  marginal shadowexon"
  11174..11254
  /note="exon predicted by xgrail, quality good"
  11690..11700
  /note="exon predicted by xgrail, quality marginal"
  join(<11814..12256,12381..12578,12676..13085,13162..13332,
  13418..>13761)
  /gene="T21P5.4"
  <11814..>13761
  /gene="T21P5.4"
  /note="similar to unknown protein GB:AAC32238 [Arabidopsis
```







```

239 sglyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProA 256
||| :|||
||| :|||
78244 TGGTTTACCTTT .....GGGATTATTACCAATGTTTCTCTA 78281

256 spSerLeuPhePheThrTyrValAsnGlnThr_SerArgSerAsnValVa 272
:|||||:|
:|||||:|
78282 CTACTCTCTTTTAGTAICTTAATCTTTAGAGTTCGAAGAAGATTT 78331
:|||||:|
272 lProValGlu ..... 275
:|:
78332 AGTAAATATATAGCTTTGCGATATAATATCTTTATAGTATCTTTGTG 78381
275 ..... 275
78382 TTTGGTTTTTCAGAAATTCAGAAACTCAAGTATTGACAGCTTTCAC 78431
276 .....AsnPhePheGlnAspAlaTyrLeuGlyValLeuProGly 288
:||||:|
:||||:|
78432 GACTACGAGCTACAAATTCAGAAAGATTGTAAAGAGGAATCTACCGAA 78481
288 spHeSerTyrIle .....AsnProSerC 296
:|:
78482 CTATGTTGCTGTCGACAACTGTCGATGATCTATTGTTGAATCCCGCT. 78530
296 yscysGlyThrAsnThrAsnSerMetHisProThrGlyAsnValSerTyr 312
:||||:|
78531 .....AACGATGATCATCCATCATGATGCTCTCGAA 78563
313 GlyGluValPheValysGlnIleTyrAspAlaIleArgGlnGlyProG1 329
:||||:|
78564 GGCNAAGCTCGTGAAGAGCTTTACGAGCCTTGAGATCGAGTCTCA 78613
329 nTrpAspIysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheT 346
:||||:|
78614 ATGGAATCAGATTTATTCTATTATAACATATACGAAACGCGTGATCT 78663
346 yAspHisValProProLeuAlaValArgProAsnLeuThrTyr 362
:||||:|
78664 AGCAGATGCTCCCTACTCCCTC .....CACGGAGTTCCTAAT 78701
363 ThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAspArgLe 379
:||||:|
78702 CCGAGTGTATCTTGGGACCTCCACCCTACAAATTTTGAATTTACCGGCT 78751
379 uGlyGlyArgMetProThrTrpValIleSerProTyrSerIysLysGly. 395
||| :|||
78752 CGGCTTAGGTTCCGACTTTCTTCATCTCCCTTGGATTGAGCTGGCA 78801
396 TyrIleGluInTyrGlyThr ..... 402
:||||:|
78802 CAGGTAAGTATGTTGGTAGACAATAATTAATTAATAATGTTGTTAAT 78851
403 .....AspProValThrGlyLysProAlap 411
:||||:|
78852 TAATGACAATTTTGTAAATGTAGTCTTCATCGGTCAAAATGGC 78901
411 ro .....TyrSerAlaThrSerValLeuLysThrLeuGly 422
:||||:|
78902 CGTACTAATGTCACATATGAGCATTCGTCAATTCGCCAACGCTCAAG 78951
423 TyrLeuTrpAspIleGluAspPhe...ThrProArgValAlaHisSerPr 438
:||||:|
78952 AAAATTTTCAACATCAAGATTTTGTGACAAAGAGAGATTCTGGGCTGG 79001
438 oSerPheAspHisLeuIle...GlyThrThrLeuArgGluAspAlaPro 453
:||||:|
79002 CACTTTTGAATCTGTTATTACGAGAAACTCACCAGAACAGATTTGCCCT 79050

```

seq\_name: gb\_ba2.MTCY98

seq\_documentation\_block:

LOCUS MTCY98 31225 bp DNA BCT 17-JUN-1998

DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 103/162.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

283860 AL123456  
283860.1 GI:3261681  
Mycobacterium tuberculosis.  
Mycobacterium tuberculosis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE  
AUTHORS

1 (bases 1 to 31225)  
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,  
Tekle, A., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
Barrell, B.G.

TITLE  
complete genome sequence  
Nature 393 (6685), 537-544 (1998)

JOURNAL  
MEDLINE  
REMARK

98295987  
Erratum: [[published erratum appears in Nature 1998 Nov  
12:396(6707):190]]

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 31225)  
Parkhill, J.  
Direct Submission  
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
75724 Paris Cedex 15, France E-mail: parkhillsanger.ac.uk  
On Jun 27, 1998 this sequence version replaced gi:1781238.

COMMENT

Notes:  
Details of M. tuberculosis sequencing at the Sanger Centre are  
available on the World Wide Web.

(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have  
been renumbered from the original cosmid submissions but the old  
gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a Hidden Markov Model of 18 genes  
implemented in TParse (Krogh) supplemented with visual inspection  
of positional base preference in codons, especially where there is  
an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct  
initiation codon. Where possible we choose an initiation codon  
(atg, gtg, or ttg) which is preceded by an upstream ribosome  
binding site sequence (optimally 5-13bp before the initiation  
codon). If this cannot be identified we choose the most upstream  
initiation codon.

FEATURES  
source

Location/Qualifiers  
1..31225  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/clone="Y98"  
<1..>31225

source

/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/clone="Y98"  
339..342  
/note="possible RBS, GCGG, for Rv2334"  
349..1281  
/gene="cysK"  
/gene="cysK"  
/note="Rv2334, (MTCY98-03), len: 310 aa. cysK, probable  
cysteine synthase. FASTA best: G540497 Cysteine synthase  
(325aa), fasta scores, opt: 1223, E(): 0, (59.5% identity  
in 309 aa overlap). Also shows good similarity to  
MTCY130.21, (323 aa), fasta scores, opt: 662, E(): 2e-30,  
(38.6% identity in 319 aa overlap) TBPASE score is 0.884;  
contains PS00901Cysteine synthase/cystathionine

RBS

gene

CDS





```

20140 .....GGTCATTTGACCGAT...ATCGAGCACATCGTGTGTCCTACAG 20099
60  GluAsnArgSerPheAspAsnLeuGlyGlyValArg...GlnG1 75
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20098  GAGAACAGGTGCTTCGATCACTATTTCGGCACGCTTCTGCGCTCGACGG 20049
75  yLeuAsnProIleAsnAsnGlyProPheCysAsnTyrLysAsnAlas 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20048  GTTCGACACTCCGACG.....CCGCTGTTTCAACAAAGGCGCTGGA 20008
92  erAspProSerSerGlyLysTyrCysThrGlnAlaLysAspTyrAspSer 108
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20007  ACCCGGAG.....ACGACGGCGCTGGACCCACCGCGC 19976
109  Val..... 109
19975  ATTACGCTGCCCTACCGCATCAATACCACCGGGGGTCCCAACGGGGTTGG 19926
110  .....PheAsnAspProAspHisSerValThrGlyAsnAsnLeuGluP 124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19925  CGAGTGGCTCAACACCCACACACAGTGGATTGGCGGCACCTTGTCAT 19876
124  heTyrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysVal 140
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19875  GG.....AACGGCGCGCC..... 19862
141  ValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLy 157
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19861  .....AATGACGGCTGGCTGCCGCGCAG.....GCGCG 19833
157  sLeuAlaProGluAlaThrArgGlnValMetGlyTyrThrGluG 174
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19832  GACCGGTGGTGGCCCAACACCGCGCTGGTGATGGGTATTACGCACGTC 19783
174  luGluValProThrLeuValAspLeuValAspGluPheThrThrPheAsn 190
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19782  CTGACATACCGATCCACTACTTGTGGCGGATCACTTCACGATCTGCGAC 19733
191  SerTrpPheSerCysValProGlyProThrAsnProAsnArgLeuCysAl 207
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19732  CAGTACTTCTCGTGGCTTGTGGCGGACGATGCTTAACCGGCTCTATTG 19683
207  aLeuAlaGlyThrAlaAlaGlyHisGly...LysAsnAspAspPheL 223
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19682  GATACGCCACCGCTCAATCCCGACGGGATCAAGTGGCGCGCAGATCG 19633
223  euAsnTyrGlyIleSerSerLysSerIlePhe..... 233
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19632  TCGAACCGGCTCCACCGCAGATTGACCTTCACCTGGCGCATATCGCG 19583
234  GluAlaAlaAsnGlyLysGlyValSerTrp..... 243
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19582  CAGAACCTCAGTCAGCGCGGCATCATGTTGGAAGGTGTACAACAGCAAGCT 19533
244  .....LeuAsnTyrAspGlyThrAsnGlyG 252
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19532  GCTCGCGGGCTCAACGACACTCTTGGCGCTACGCGGTATGCGGCA 19483
252  luPhe.....GluProAspSerLeuPhePheThrTyrValAsn 264
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19482  GTTTCMAACAGCGCGCAGATCCGAGGTG..... 19454
265  GlnThrSerArgSerAsnValValProValGlu.....AsnPhePheG1 279
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19453  GACCTGGCGGCTTATGGCATCGCCCGGCTACCGGTGGATTTCATCCG 19404
279  nAspAlaTyrLeuGlyValLeuProLysPheSerTyrIleAsnProSerC 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19403  CGAGCTCATCAACACACGCTGCCCGAGGTGCTCGGTGGTTCG..... 19358
296  yscysGlyThrAsnThrAsnSerMetHisProThrGlyAsnValSerTyr 312
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19357  .....TTGACCGTGGAGTCCGACATCCGTCATCCGCTGGCGCATC 19316
313  GlyGluValPheValLysGlnIleTyrAspAlaIleArgGlnGlyProG1 329
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19315  GGTGGGTGACGATCGTGAACATTGATAAGGTGTTGCTGCGCAATCCGC 19266
329  n....TrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyP 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19265  GGTGGGAGAAACCGCTTGATCATCGCTATGACGAACATGCGCGCT 19216
345  heTyrAspHisValProProLeuAlaValArgProAspAsnLeuThr 361
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19215  TCTTCGACACGCTACACCGCTCACCGCG.....CCGGAGGCG...ACA 19175
362  TyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAspAr 378
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19174  CCCGCGAATGGATTCCCAAC.....AGTGTGACATCGACAA 19137
378  g.....LeuGlyGlyArgM 383
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19136  GGTGACGCGCTCCGCGGAATACGTGGACCATCGCTTGGGCTTCGCG 19087
383  etProThrTrpValIleSerProTyrSerLysLysGlyTyrIleGluGln 399
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19086  TGCCCTGCTTCGTCATTTCGCTTACAGTCGCGGCGGCTGATGTCAT 19037
400  TyrGlyThrAspProValThr.....G1 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19036  GATCGGTTCCGACACACATCGCAGCTGCAATTGATCGCAAGCGTTTCGG 18987
407  yLysProAlaProTyrSerAlaThrSerValLeuLysThrLeuGlyTyrL 424
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19896  GTGCGGTTCCCACTTGCACACTGCACCTGGCTGCCAGTGTACCGCGGATA 18937
424  eutrAspIleGluAspPhe...ThrProArgValAlaHisSerProSer 439
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18936  TGACGTGGCATCAATTTCGCGGCGCGCGGACCGCTGCGCACCAAT 18887
440  PheAspHis..... 442
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18886  CTGGACCAACCGGTCGCTCAATTGCCGAAGTCCCAAGTCCGTCGCCAA 18837
443  LeuIleGlyThrLeuArgGluAspAlaProIleAlaLeuLysThrP 459
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18836  TGTGTGTGTGGTTTCTTGACGAAGGCTGCGGTATCGGGTGCCCTACC 18787
459  roHisThrPheSerVal 464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18786  CCCAACACCGCCAGTC 18770
seq_name: gb_ba1:AF107252
seq_documentation_block:
LOCUS AF107252 2262 bp DNA BCT 14-DEC-1999
DEFINITION Burkholderia pseudomallei non-hemolytic phospholipase C gene,
complete cds.
ACCESSION AF107252
VERSION AF107252.1 GI:5572520
SOURCE Burkholderia pseudomallei.
ORGANISM Burkholderia pseudomallei.
Bacteria; proteobacteria; beta subdivision; Burkholderia group;
Burkholderia; pseudomallei group.
1 (bases 1 to 2262)
Korrisasate,S., Suwanasai,N., Leelaporn,A., Ezaki,T., Kawamura,Y.
and Sarasombath,S.
Cloning and characterization of a nonhemolytic phospholipase C gene
from Burkholderia pseudomallei
J. Clin. Microbiol. 37 (11), 3742-3745 (1999)
MEDLINE 99454884
PUBMED 10523590
2 (bases 1 to 2262)
Korrisasate,S., Suwanasai,N., Leelaporn,A., Kawamura,Y., Ezaki,T.
and Sarasombath,S.

```





```

/transl_table=11
/product="RvD2-Rv1758"
/protein_id="CAB44660.1"
/db_xref="GI:5042243"
/translation="MVTWALLFAPVPAASADPPDPTVSDGACPDVEYVFARCTGPPPP
GVGGIGEDFIDALRSKIGKSMGVYGVDPATTFDPTAMAGIYDAGTHVEQTAAACPCQ
SKLVGGFSQGAAMVGEVTAALIPDGAFLDAPRMPPEVAD"
misc_feature
7657..7937
/gene="RvD2-Rv1758"
/notes="corresponds to M. tuberculosis complete genome
region 1989055..1989337, AccNo AL123456"
BASE COUNT      1420 a  2422 c  2599 g  1496 t
ORIGIN

```

```

alignment_scores:
  Quality: 308.00      Length: 443
  Ratio: 1.474        Gaps: 20
  Percent Similarity: 47.178      Percent Identity: 27.991

```

## alignment\_block:

US-09-426-072-2 x MBOY18606/rev ..

Align seg 1/1 to reverse of: MBOY18606 from: 1 to: 7937

```

9 GlyLeuGlyLeuAlaSerLeuGlyLeuAlaSerProValThrSerG1 25
|||||
1484 GGCCTGGCGCGCTGATGCTCTCTTCGGCGGTCGGTGATT.....1446
25 uTyrThrSerValArgGluAlaProPheGlyTyrLysProGlySerLysG 42
|||||
1445 .....GAAAAGGCTACGGCGGGGCGCTGCTCC....1416
42 LuSerIleGluAsnLeuLysAspLysValGluAsnIleValTrpLeuIle 58
|||||
1415 .....GGCATTATGACGAC...ATCGAACACTTGTGTTCTTCATG 1377
59 LeuGluAsnArgSerPheAspAsnIleLeuGlyGlyValArgArgGlnG1 75
|||||
1376 CAGGAGAACCGCTCATTCGATCACTACTTCGGAACACTC.....TCCGG 1333
75 yLeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrLysAsnAla 92
|||||
1332 CACGATGATTCAACACGGTATCGCGCTCTTCCAAAGAGGGTTGA 1283
92 erAspProSerSerGlyLysTyrCysThrGlnAlaLysAspTyrAspSer 108
|||
1282 ACCCG.....ATGACGAGGCTCTTGACGCCACCGGC 1251
109 Val.....109
1250 GTTACGATCCCTACCGTTCGACACACCGAGCGCGTTCCTCCACGG 1201
110 .....PheAsnAspProAspHisSerValThrGlyAsnAsnLeuGluP 124
|||||
1200 CGCGTGTGTCAACGACCGCTACCATAGC.....T 1172
124 heTyrGlyThrThrProAsnAsnGlyAlaIleAlaSerGlyLysVal 140
|||
1171 GGGTGGGATGATGATGCTGGAAATGGTGGCGTC.....1137
141 ValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLy 157
|||||
1136 .....AAGCAACACTGCTACCGAGCGCAG.....GCCAA 1108
157 sLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrThrGluG 174
|||||
1107 GACTCGCTGGCGGGCCACACTCCAGGTGATGGCTACTACACGGCC 1058
174 luGluValProThrLeuValAspLeuValAspGluPheThrPheAsn 190
|||||
1057 AAGACATTCGATCCACTATCTGTGGCGGACCGCTTACCGCTCTCGAC 1008

```

```

191 SerTrpPheSerCysValProGlyProThrAsnProAsnArgLeuCysAl 207
|||||
1007 CGCTACTTTTGTTCGGTATTGGGTGCCACCTTGCCTCCCAACCGCTTTATTG 958
207 aLeuAlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeu. 223
|||||
957 GCTGAGCGCCACCATCGACCCCGACGAGCCCAAAATGGTGGCGGAGCTGC 908
224 .....AsnTyrGlyIleSerSer 229
907 AGAGTCCCACTTCCACACCGGTCCGAGATTCGGTGGCGCATCATGCGC 858
230 LysSerIlePheGluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTy 246
|||||
857 CAGAACCTTAGCGATGCC.....GGCGTCAGCTGGAAGGTGTA 820
246 rAspGlyThrAsnGlyGluPheGluProAspSerLeuPhePheThrTyV 263
|||||
819 T.....CGCAACAGACACTCGGCGCCCATCTCTCGGTCTTACTTAGC 776
263 aAsnGlnThrSer.....267
775 GCTCGCTTGTGACGCTTTTCAACAGTCAGCGATCCAGGTCAGATCTT 726
268 ...ArgSerAsnValValPro.....ValGluAsnPhePheGlnAspAl 281
|||||
725 GTCGGCTTGGCGTGGCACCAGCTATCCCGGAGCTTTCGGCGCCGACGT 676
281 aTyrLeuGlyValLeuProLysPheSerTyrIleAsnProSerCysCysG 298
|||||
675 CTTAGCAATAGACTCGCGGGGTCTCTGGGTGATTCCCAATGTT....630
298 lyThrAsnThrAsnSerMetHisProThrGlyAsnValSerTyrGlyGlu 314
|||||
629 .....CTCGAATCCCAACATCTTCCGGTTCAGCGCGCGCGCGGCT 588
315 ValPheValLysGlnIleTyrAspAlaIleArgGlnGlyProGln...Tr 330
|||||
587 TTCGAATGTCACATCTTAAGAATATTGCTTCCCAATCTCGCGGTGTG 538
330 pAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheTyA 347
|||||
537 GGAAGAGCGCGCTGATCGTCAGCTACGACGAAACGGCGCTTTTCG 488
347 sphisValProProLeuAlaValArgProAspAsnLeuThrTyrThr 363
|||||
487 ACCAGCTTCTCTCTACCGCG.....465
364 GluThrAlaLysAsnGlyGlnLysTyrThrLeu...HisPheAspArg... 378
|||||
464 CCGCGCGGACTCCCGCGGAATATGTCAGCGTCCGCTGACATCGATCATG 415
379 .....LeuGlyGlyArgMetP 384
414 GCGCGGCTCCGCGGAATACGCGGCGGATCGGTTCGGCTTCGGCTTC 365
384 roThrTrpValIleSerProTyrSerLys 393
|||||
364 CTGCTCTGCTCATTTCCGCGGTACAGCGCT 336

```

seq\_name: gb\_p11:AC005168

seq\_documentation\_block:

LOCUS AC005168 91893 bp DNA PLN 05-APR-2000

DEFINITION Arabidopsis thaliana chromosome II section 150 of 255 of the complete sequence. Sequence from clones F10A8, F12C20, T20P8.

ACCESSION AC005168 AE002093

VERSION AC005168.2 GI:6598444

KEYWORDS HTG.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;













```

4503 CGTCACTGCTGTACCTGTACGTGACGCGCTACTGGAAATACCGGCTCCAG 4454
273 ProValGluAspPhe..... 277
4453 GAGCAGAACTACGACTGCAAGCGCCCTCGCTGGTTCAGGAACATCA 4404
278 .....PheGlnAspAlaTyrLeu..... 283
4403 CAACGCTCCGCGCGATTCGACCTCTGGCAGCGCCCATGCTCGCCGCG 4354
284 .....GlyValLeuProLysPhe 289
4353 CGCTGACAGCTGGCGAAGGATGTCGAGGAGAACACGCTCGCGGAGTG 4304
280 SerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMetHisPr 306
4303 TCGTGGATGCTCGCGCCCTACTGCTACTCGGAG.....CATCC 4266
306 oThrGlyAsnValSerTyrGlyGluValPheValLysGlnIleTyrAspA 323
4265 CTGCTGGGCGCCCTCGTTCGGCGGAGTACTACGTGACCGGGTCTCGAG 4216
323 lalLeuArgGlnGlyProGln...TrpAspLysThrLeuLeuPheIleThr 338
4215 CGCTGACAGCAACCCGAGGTCTGGCCGACGACCGTGTTCATCTCAAC 4166
339 TyrAspGluThrGlyGlyPheTyrAspHisValProProProLeuAlaVa 355
4165 TACGACGAGGCGCGGCTCTATGACACGCGCCGCGCGGTCGCGCC 4116
355 lArgProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysT 372
4115 GTGCAAGGATGCGTCCGCGCTTCCACGCTCAGCACCGCGCGAG.... 4070
372 yrThrLeuHisPheAspArg.....LeuGlyGlyArgMet 383
4069 .....ATCAGAGTATCCAGCGGGTCCCATCGGCTCGGCCACCGCGTC 4025
384 ProThrTrpValIleSerProTyrSerLysLysGlyTyrIleGluGlnTyr 400
4024 CCGTGATCGCCATCTCGCCCTGTGCGAAGCGCGCAAGGTC..... 3983
400 rGlyThrAspProValThrGlyLysProAlaProTyrSerAlaThrSerV 417
3982 .....AGCGCCGCAAGTGTTCACCATCATCCCTCG 3955
417 alLeuLysThrLeu..... 421
3954 TGCTGCGCTTCTCGAAGCGCGTTCGCGGCTCGTGAGGAAACATCTCG 3905
422 .....GlyTyrLeuTrpAspIleGluAspPhe... 430
3904 CCGTGGCGCGTGGCGTCTGCGGCGACCTGACTTGGCTGTTCGACTTCCA 3855
431 .....ThrProArgValAlaHisSerProSerP 440
3854 GGGCGGCGGATACCCAGGTGCGCCCGACCTGACCAACGTCGCCGACA 3805
440 heAspHisLeuIleGlyThrThrLeuArgGluAspAla 452
3804 GCGAC.....GCGCGCAGGAGGAGCGCC 3782
seq_name: gb_ba2:MSGMTP40A
seq_documentation_block: 3013 bp DNA circular BCT 16-JUN-1994
LOCUS MSGMTP40A
DEFINITION Mycobacterium tuberculosis phospholipase C gene and MTP40 protein, complete cds.
ACCESSION L11868
VERSION L11868.1 GI:399585
KEYWORDS MTP40 protein; phospholipase C.
SOURCE Mycobacterium tuberculosis (strain H37Rv) (library: sub-library of BAM HI M. tuberculosis DNA).

```

```

ORGANISM
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
AUTHORS
1 (bases 1 to 3013)
TITLE
Parra,C.A., Londono,L.P., del Portillo,P. and Patarroyo,M.E.
Isolation, characterization, and molecular cloning of a specific
Mycobacterium tuberculosis antigen gene: identification of a
species-specific sequence
Infect. Immun. 59, 3411-3417 (1991)
91372944
JOURNAL
MEDLINE
2 (bases 1 to 3013)
REFERENCE
AUTHORS
del Portillo,P., Murillo,L.A. and Patarroyo,M.E.
TITLE
Amplification of a species-specific DNA fragment of Mycobacterium
tuberculosis and its possible use in diagnosis
J. Clin. Microbiol. 29, 2163-2168 (1991)
92042637
JOURNAL
MEDLINE
3 (bases 1 to 3013)
REFERENCE
AUTHORS
Leao,S.C., Rocha,C.L., Parra,C.A. and Patarroyo,M.E.
TITLE
Nucleotide sequencing of a 3.0 kbp fragment of Mycobacterium
tuberculosis DNA
Unpublished (1993)
JOURNAL
FEATURES
Location/Qualifiers
1..3013
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/tissue_lib="sub-library of BAM HI M. tuberculosis DNA"
462..2000
/citation={3}
/codon_start=1
/transl_table=1
/label=M-TBC-EXCLUSIVE
/product="phospholipase C"
/protein_id="AAB59164.1"
/db_xref="GI:399586"
/translation="MSRREFTLKLGTGAGAAFLMDWAAPVIEKAYGAGPCGHLTDIE
HIVLQENRSFDHYFGTSLSTGFGNAASPAFOGWNPMTOALDPAGVTIPERLDTT
RGPFIDGCVNDPEHVMGMLANGSANDMLPAQATTRAGPYPLMGVYTRQDTP
IHYLLADFTICDGYHCSLTGTLPNRLYMLSANIDAGTDGGPOLYVERGFLPQQFS
WRIMPENLEDAGSVKVKYQKNGLCGRFINTPISNGLVQAFRQADPRNLARYGIAPT
YQGFADVRANRPKPSVNLQSEHPALPVALGAVSMVTALRLLSNPAWVEKT
ALIVSYDNGFFDHYTPPTAPGTEFVTPNIDAVPGSGGIRGLGFRVPCIV
ISPYSRGLMSVDTEDHTSQLIRAREGVVPVNTAMRDGVYDMTSAFNATPPNS
TRPNLSHPLLKALPKLPQCIPNVVLGTTDGLPSTPYRVYPYQVMPETOETTPVRGTPS
GLCS"
old_sequence
complement(1417)
/citation={1}
/citation={2}
/replaces="a"
complement(1478..1915)
/citation={1}
/citation={2}
/codon_start=1
/transl_table=1
/label=W-TBC-EXCLUSIVE
/product="MTP40 PROTEIN"
/protein_id="AAB59165.1"
/db_xref="GI:399587"
/translation="MLGNAPSVVPTNTLGMHCGSFGSAPSNGMLKGLVFERGVAKLN
AEVMSPTTSRQAVMLGTGTPNRAINFNCEWNSVSETISGPRLYGEMTWQGTGRKPR
PSGPRMPDPGTAASMLGTNTNSPGVGGAVGVTSKPPFS"
old_sequence
complement(1530..1532)
/citation={1}
/citation={2}
/replaces="cc"
complement(1940..1941)
/citation={1}
/citation={2}
/replaces="g"
complement(1980..1982)
/citation={1}
/citation={2}

```

```
BASE COUNT      585 a      /replace="cc"      883 g      558 t
ORIGIN          987 c
```

alignment_scores:		
Quality	286.00	Length: 447
Ratio	1.388	Gaps: 20
Percent Similarity	46.085	Percent Identity: 28.635

```
alignment_block:
US-09-426-072-2 x MSGMTP40A ..
```

Align seg 1/1 to: MSGMTP40A from: 1 to: 3013

```

7  LeuThrGlyLeuGlyLeuLeuAlaSerLeu...GlyLeuAlaSerProVa 22
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
489  CTCACTCGCGCAGCGCAGCGCATCTCTGATGAGCACTGGCGCTGCACCGCT 538
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
22  lThrSerGlyThrThrSerValArgGluAlaProPheGlyThrLysProG 39
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
539  GATTGAAAAGCCCTACGCCCC.....GGCCTTTGCTCCCG 573
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
39  lysertylsuSerIleGluAsnLeuLysValGluAsnIleVal 55
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
574  GA.....CATTTGACCGAC.....ATCGAGCATATCTGTG 602
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
56  TrpLeuIleLeuGluAsnArgSerPheAspAsnIleLeuGlyGlyValar 72
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
603  TTGCTCATGACGGAGAACCGCGTCATTTCGACCACTATTTCGAGACGCTTC 652
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
72  gArgGlnGlyLeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrL 89
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
653  CAGC.....ACCAATGGGT 666
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89  ysAsnAlaSerAspProSerSerGlyLysTyr.....CysThr 101
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
667  TCAACGCGCGTCGCGCGCATTCACAAATAGGGTTGGAACCCCATGACG 716
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102  GlnAlaLysAspTyrAspSerVal..... 109
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
717  CAGCGCTTGACCCCGCGCGGTGCACATTCGCTTCGCTTGACACCA 766
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110  .....PheAsnAspProAspHisSerV 117
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
767  CCGAGCGCCCTTCCTGACGCGCGAGTGCCTCAACGACCCCGCAGCAGCT 816
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117  alThrGlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAsnAsnGly 133
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
817  GGTGGGGATGCACCTGGCCCTGG.....AACGGTGTG 848
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134  AlaIleAlaSerGlyLysValAlaAlaAspGlnSerGlyPheLeuAsnAl 150
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849  GCC..... 851
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
150  aGlnLeuAsnAspTyrProLysLeuAlaProGluAlaThrArgGln. 166
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
852  .....AACGAC.....AAGTGGTCGCGCGCAGCGACCCCGCGG 888
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167  .....ValMetGlyTyrTyrThrGluGluVal 176
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
889  CAGGACCATATGTCCTTTGACCATGGGTACTACAGCGGCCAAGACATC 938
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
177  ProThrLeuValAspLeuValAspGluPheThrThrPheAsnSerThrP 193
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
939  CCATCCATCATCTGCTGGCGGACAGCTTCACCATCTGCGACGGCTACCA 988
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
193  eSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeu... 208
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
989  TTGCTGCTGTCAGCGGACCGCTGCCAACCGCGCTACTGCTGTTCAGCG 1038
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209  .....AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAsp 220
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

1039 CCAACATCGACCCCGCGGCACCGACGGGGAGACCCCAATTGGTAGAGCGC 1088

221 AspPheLeuAsnTyrGlyIleSerSerLysSerIlePhe...GLuAlaAl 236  
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1089 GGGTCTCTCGCGCTGCAGCAATTGAGTTGGCGCATCTGCGCGGAAACCT 1138

236 aAsnGlyLysGlyValSerTrpLeuLeuAsnTyrAspGlyThrAsn...GlyG 252  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1139 CGAAGATCCGGGTGACGTGGAGAGTGTACAGAACAGAGCGCTTCGGCG 1188

252 LuPheGluProAspSerLeuPhePheThrTyrValAsnGlnThrSerArg 268  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1189 GATTTCACAACGCCCATCAGCAATAACGGCGTGGTCAGCGCTTCGCC 1238

269 SerAsnValProValGluAsn..... 276  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1239 CAGCGCATGATCCGAGGTGCAAGTCTGGCCGCTACGGTATCGCCCGCAC 1288

277 ..... PhePheGlnAspAlaTyrLeuGlyValLeuProLysP 289  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1289 CTACCTCGGGGACTTCCTCGCGAGTCAGGGCCACCGGCTACCCACAG 1338

289 heSerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMetHis 305  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1339 TCTCTGGTTAGTTCCCAACATC.....CTGCATCCGACAC 1376

306 ProThrGlyAsnValSerTyrGlyLuValPheValLysGlnIleTyrAs 322  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1377 CCGCCCTCGCGTAGCGTCTGGCCGGTGCTCCACGGTGACCGCGCTCGC 1426

322 pAlaIleArgGlnGlyProGln...TrpAspLysThrLeuLeuPheIleT 338  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1427 GATCTTGTGTGTCCAATCCCGCGGTGTGGAAAGACCGCATTTATGCTCA 1476

338 hrTyrAspGluThrGlyGlyPheTyrAspHisValProProLeuAla 354  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1477 GCTATCAGAGAACGGCGGCTCTCTCGACCATCGTCACGCCCCCGCAGCCA 1526

355 .....ValArgProAspAsnLeuThrTyrThr 363  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1527 CCGCCCGGGACACCCGGCGAAATCGTCACGGTCCCAACATCAC...GC 1573

363 rGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAspArgLeuG 380  
: : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

1574 AGTACCCGGGTCCGGTGGATCTGGTGGTCCGCTC.....GGTCTGG 1614

380 LyGlyArgMetProThrTrpValIleSerProTyrSerLys 393

1615 GTTTTCGGCTTCCTGCATTGTTCATTTCGCGCTACAGCGCG 1655

seq\_name: qb\_ba2:MTU49511

seq\_documentation\_block: 4094 bp DNA BCT 05-JUN-1998  
LOCUS MTU49511  
DEFINITION Mycobacterium tuberculosis phospholipase C (mpcA) and phospholipase C (mpcB) genes, complete cds.  
ACCESSION U49511  
VERSION U49511.1 GI:1399492  
KEYWORDS  
SOURCE Mycobacterium tuberculosis.  
ORGANISM Mycobacterium tuberculosis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE  
1 (bases 1 to 4094)  
Mycopacterium, Mycopacterium

**AUTHORS** Johansen, K. A.

**TITLE** Direct Submission

JOURNAL Submitted (20-FEB-2019)

Microbiology, University of...  
Ninth Ave. ...  
...

FEATURES

FEATURES	LOCATION
source	1. .4094

source	1. .4034
..	/organism

1

\_\_\_\_\_



```

1139 CGAAGATCCCGGGGTGAGTGGAGGTGTACCAAGAGGAGGCGCTCCGGC 1188
252 luphegluProAspSerLeuPhePheThrTyrValAsnGlnThrSerArg 268
::||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1189 GATTTCATCAACACGCCCATCAGCAATAACGGGTGTGGTCCGCTCCGC 1238
269 SerAsnValValProValGluAsn..... 276
::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1239 CAGGCAGCTGATCCAGGTGCAATCTGGCCCGTACGGTATCGCCCGCAG 1288
277 .....PhePheGlnAspAlaTyrLeuGlyValLeuProLysP 289
||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1289 CTACCTGGGAGCTTCGCTGCCAGCTCAGGCCAACCGCTATCCCAAG 1338
289 heSerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMetHis 305
||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1339 TCTCTGTGTAGTTCCTCCCAATC.....CTGCAGTCCGAACAC 1376
306 ProThrGlyAsnValSerTyrGlyValPheValLysGlnIleTyrAs 322
||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1377 CCGCGCTGCGGTAGCTGTGGCGGTGTCCATGTGACCGCGCTGG 1426
322 palalleArgGlnGlyProGln...TrpAspLysThrLeuPheIle 338
::| ::| ::| ::| ::| ::| ::| ::| ::|
1427 GATCTGTGTGTCAATCCCGCGGTGTGGAAAGACCGCTATTCGTCA 1476
338 hrTyrAspGluThrGlyPheTyrAspHisValProProLeuAla 354
::| ::| ::| ::| ::| ::| ::| ::| ::|
1477 GCTATGACGAGAACGGCGCTTCTTCGACACCGCTACGCGCCCGCAG 1526
355 .....ValArgProAspAsnLeuThrTyrTh 363
||| ::| ::| ::| ::| ::| ::| ::| ::|
1527 CCGCGCGGACACCGCGCAATCTGACGGTGCCCAATCGAC...GC 1573
363 rGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAspArgLeu 380
::| ::| ::| ::| ::| ::| ::| ::| ::|
1574 ACTACCGCGGTGCGGTGGCATTCGTGGTCCGCTC.....GGTCTGG 1614
380 lylGlyArgMetProThrTrpValIleSerProTyrSerLys 393
||| ::| ::| ::| ::| ::| ::| ::| ::|
1615 GTTTTCGGTTCCTGCATGTCATTCGCGGTACAGCGC 1655

```

seq\_name: gb\_ba2:PSEPLCN

```

seq_documentation_block: 2932 bp DNA BCT 26-Apr-1993
LOCUS PSEPLCN
DEFINITION P.aeruginosa nonhemolytic phospholipase C gene, complete cds.
ACCESSION M59304
VERSION M59304.1 GI:151495
KEYWORDS phospholipase C.
SOURCE P.aeruginosa DNA.
ORGANISM Pseudomonas aeruginosa
Bacteria; proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.

```

```

REFERENCE
1 (bases 1 to 2932)
AUTHORS Ostroff R.M., Vasil A.I. and Vasil M.L.
TITLE Molecular comparison of a nonhemolytic and a hemolytic
phospholipase C from Pseudomonas aeruginosa
JOURNAL J. Bacteriol. 172, 5915-5923 (1990)
MEDLINE 91008968
FEATURES
source
1..2932
/organism="Pseudomonas aeruginosa"
/db_xref="taxon:287"
gene 759..2837
/gene="plcn"
CDS 759..2837
/gene="plcn"
/note="nonhemolytic phospholipase C"
/codon_start=1
/transl_table=11
/product="phospholipase C"
/protein_id="AA25968.1"

```

```

/db_xref="GI:151496"
/translation="MISKRRSRFIRLAAGTAAATVATSMPLSPSSIOAALAIPARRHGN
LKDVHVVILMQENRSFHYFTGLKVGFRGDMALPLDQQRVWHQKSGKEILPYH
FDTSTTSAQVDTGTPHTWPAQQAQWGRMDKWLPAKTERSLGYKYEODIAFOFAMAN
AFTICDAYHCSFGQGTNPNRLFTWTNDPLGQGGPVTTNDHDSNGPVEQGYTWTTY
PERLQAAGLTHRWYODMADNPSNPLIGFROYRAAPDPSPLVNGLSWKLDALKRDV
LANSLPQSWIYAPAKYSEHPGSSPIWGAETYSWLDALTANPEVWSKTALLVWFDE
NGSFDHVAAPASPLNDKGLTKTADATLEHTKGDRIYRNQPYGLGARVPYVI
SPGSKGWNVSQVDFHTSVIRLEQRFYMEPNISPMWRRAVCGDLTSAEFNANPNNEP
FPELPTSQADAIVASQIKLPKPKPAAMPKQEMGIRPARALPYELGVHARYRSQG
DALSLTANTKAGAVQVFDLDSNPPKRYTGVGARRKRLDSFGQDASGDYHLEVHG
PNGFLRVFNGNLRDLADGKAPLEVRIDYELPLGNLRVOLINRGRRHPVKLVKDNVY
ROGERRTVANPQQRREVRYSLRSGNWYDFSVSAQGADSLFRFSRMEGDSGFSFD
PGMGLGTLTF"

```

```

BASE COUNT 505 a 1000 c 957 g 470 t
ORIGIN

```

```

alignment_scores:
Quality: 274.50 Length: 497
Ratio: 1.139 Gaps: 21
Percent Similarity: 48.491 Percent Identity: 24.950
alignment_block:
US-09-426-072-2 x PSEPLCN ..
Align seg 1/1 to: PSEPLCN from: 1 to: 2932

```

```

11 GlyLeuLeuAlaSerLeuGlyLeuAlaSerProValThrSerGluTyrTh 27
||| ::| ::| ::| ::| ::| ::| ::| ::|
801 GGTACGTCGCGCGCACCGCTGCCACCGACGATCTCCGCTCGACATCCA 850
27 sSerValArgGluAlaProPheGlyTyrLysProGlySerLysGluSerI 44
::| ::| ::| ::| ::| ::| ::| ::|
851 GCGCGCGCTCGCGATTCGCGCGCACCGCGCGATGCGC..... 887
44 lGluAsnLeuLysAspLysValGluAsnIleValTrpLeuIleLeuGlu 60
||| ::| ::| ::| ::| ::| ::| ::| ::|
888 .....AACCTCAAGGAC...GTCGAGCAGCGTGGTGTCTGTATCGAG 929
61 AsnArgSerPheAspAsnIleLeuGly.....GlyValArgArgG 74
||| ::| ::| ::| ::| ::| ::| ::| ::|
930 AACCGCTCTTCACCATCTATTTCGCGCACCCCTCAAGGGCGTCCGCGCTT 979
74 nGly.....LeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrL 89
||| ::| ::| ::| ::| ::| ::| ::| ::|
980 CGCGCAGCGCATGGCGATCCCGCTGCCGATGGCCAGCGGTCTGGCACC 1029
89 ysAsnAlaSerAspProSerSerGlyLysTyr.....Cys 100
::| ::| ::| ::| ::| ::| ::| ::|
1030 AGAAGGCGCAAGGCGCGATCTCTGCCCTACCATCTCGACACCGACACC 1079
101 ThrGlnAlaLysAspTyrAspSerValPheAsnAspProAspHisSerVa 117
||| ::| ::| ::| ::| ::| ::| ::| ::|
1080 ACCAGCGCGCGCGCTCGACGCGC..... 1103
117 lThrGlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAsnAsnGlyA 134
||| ::| ::| ::| ::| ::| ::| ::| ::|
1104 .....ACCCGCGACACACCTGGC 1119
134 lAlleAlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAla 150
||| ::| ::| ::| ::| ::| ::| ::| ::|
1120 CGGACGCG.....CAGCAGCGCTGGAGCAAGGG 1148
151 GlnLeuAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnVa 167
::| ::| ::| ::| ::| ::| ::| ::|
1149 CGCATGTGAC.....AAGTGGCTCCGCGCAAG...ACCGAGCGCTTC 1186
167 lMetGlyTyrTyrThrGluGluGluValProThrLeuValAspLeuVala 184
::| ::| ::| ::| ::| ::| ::| ::|
1187 CTGGGCTACTACAAGGACGAGGACATCGCTTCCAGTTCGCGATGGCCA 1236
184 spGluPheThrThrPheAsnSerTrpPheSerCysValProGlyProThr 200

```



```

888 .....AACCTCAAGGAC....GTCGAGCACGTGGTGTCTGATCTGATCGAGGAG 929
61  AsnArgSerPheAspAsnIleLeuGly.....GlyValArgArgG1 74
|||||
930  AACCGCTCTCTTCGACCACTATTTCGGCACCTCAAGGGCGTCCGGCGCTT 979
74  nGly.....LeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrL 89
|||||
980  CGCGCCGCGATCGGATCCCGCTGCGGATGCCAGCGGGTCTGGCACC 1029
89  yAsnAlaSerAspProSerSerGlyIlyTyr.....Cys 100
|||||
1030  AGAGGGCAGCAAGGGCGAGATCCTGCCCTACCACTTCGACACCAAGCACC 1079
101  ThrGlnAlaLysAspTyrAspSerValPheAsnAspProAspHisSerVa 117
|||||
1080  ACCAGCGCCAGCGCGTCGACGCG..... 1103
117  lThrGlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAsnAsnGlyA 134
|||||
1104 .....ACCCGCGCACACCTGGC 1119
134  laIleAlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAla 150
|||||
1120  CGGACGCC.....CAGCAGCGCTGGAGCAAGGG 1148
151  GlnLeuAsnAspTyrProLysLeuAlaProGluAlaAlaThrArgGlnVa 167
|||||
1149  CGCATGGAC.....AAGTGGTGGCGGCCAAG...ACGAGCGTTC 1186
167  lMetGlyTyrThrGluGluValProThrLeuValAspLeuValA 184
|||||
1187  CTGGGCTACTACAGGAGGACGACATCGCCTCCAGTTCGGCGATGGCCA 1236
184  spGluPheThrPheAsnSerTrpPheSerCysValProGlyProThr 200
|||||
1237  ACGCCCTTACCATTCTGGAGCGCTATCACTGCTGCTCCAGCGCGGCACC 1286
201  AsnProAsnArgLeuCysAlaLeuAlaGlyThr.....AlaAlaG1 214
|||||
1287  AACCCCAACCGGCTGTTCTCTGACCGCGCACCAACGACCGCTCGGCCA 1336
214  yHisGly.....LysAsnAspAspPhe.....LeuA 224
|||||
1337  GCACGGTGTGCGGTACCAACCAACGACGACGACGACGACGCGCGGTGG 1386
224  snTyrGlyLeuSerSerLysSerIlePheGluAlaAlaAsnGluLysGly 240
|||||
1387  AGCAGGCTACACCTGGACCACTATCCCGAGCGCTCGAGCTGCCGCGC 1436
241  ValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAspSe 257
|||||
1437  ATCACCCTGGCGGGTCTACCAAGGACATGGCCGACAACTTCCGACCAACC 1486
257  rLeu.....PhePheThrTyrValAsnGlnThrSerArgSerAsnValV 272
|||||
1487  CTGTATCGCTTCGCCAGTACCGCGCGCGCGCCCTGACTCGCGCTGA 1536
272  al.....ProValGluAsnPhePheGlnAspAla 281
|||||
1537  TCGTCAACGGCTCGAGCACCTGGAGCTCGATGCGCTGAAGCGGACGCTG 1586
282  TyrLeuGlyValLeuProLysPheSerTyrIleAsnProSerCysCysG1 298
|||||
1587  CTGGCAATAGCTTCGCCAGGTCCTGGATC.....GTGCG 1624
298  yThrAsnThrAsnSerMetHisPro...ThrGlyAsnValSerTyrGlyG 314
|||||
1625  CCGCGGCAAGTACTCCGAACACCGCGCGCTCCAGCCCGATCTGGGGG 1674
314  luValPheValLysGlnIleTyrAspAlaIleArgGlnGlyProGln... 329
|||||
1675  CCGAGTACACTTCTGGGTACTCGAGCGGCTGACCGCAACCGGAGGTCT 1724

```

```

330  TrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheTy 346
|||||
1725  TCGAGCAAGACCGCGCTGCTGTGATGTTTCGACGAACGACGCGTTCIT 1774
346  rAspHisValProProProLeuAla.....ValArgProAspAsnLeuT 361
|||||
1775  CGACACAGTCCGCCCGCGCGCGCGCGCTGACACAGGACGCGCACGC 1824
361  hrTyrThrGluThr.....AlaLysAsn 368
|||||
1825  TCGCTGCAAGACCAACCGCGCGACGCCACCTCGAATGTCACACCAAGGC 1874
369  GlyGlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProTh 385
|||||
1875  GATATCGTTATCGCAACGACGCCCTACGGCTCGCGCGCGGTCGCGAT 1924
385  rTrpValIleSerProTyrSerLysLysGlyTyrIleGluGlnTyrGlyT 402
|||||
1925  GTACGTGATCTCGCGTGGAGCAAGGCGCTGGGTCAACTCCCA.GGTG 1973
402  hrAspProValThrGlyLysProAlaProTyrSerAlaThrSerValLeu 418
|||||
1974  TTCGACCAACACCTCGCTGATCGCTTCTCGAGCAGCGCTTCGG..... 2017
419  LysThrLeuGlyTyrLeuTrpAspIleGluAspPheThrProArgValAl 435
2018 .....GG 2019
435  ahISerProSerPheAspHisLeuIleGlyThrThrLeuArgGlu...A 451
|||||
2020  TCATGGAGCGAATATCAGTCCCTGGCGTCTGCTGCGCGGACCTG 2069
451  spAlaProIleAlaLeuLysThrProHisThrPheSerVal 464
|||||
2070  ACCTCGCGTCAACTTCGCCCAACCGCAACACGAGCGCGTT 2110
seq_name: gb_bal:AE004754
seq_documentation_block:
LOCUS AE004754 10092 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 315 of the complete
genome.
ACCESSION AE004754 AE004091
VERSION AE004754.1 GI:9949446
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 10092)
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,
Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y.,
Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
NATURE 406 (6799), 959-964 (2000)
JOURNAL
MEDLINE
2043737
REFERENCE
2 (bases 1 to 10092)
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,
Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H.,
Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
Location/Qualifiers

```



```

source      1. .10092
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
93. .821
/gene="PA3317"
93. .821
/gene="PA3317"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06705.1"
/db_xref="GI:9949447"
/translation="MRKPLLSALLGLGALGCSLPQOQAGAPTEQRIGTQMGEGVAS
PVTVALRLSQPVDRQVYFASRFDGAIKELPLAKRGVGFVLDDEGKFDLVO
HRTTLOQREGQRYRLWLNANLSATYEVVATVDGLVNGQPSLKNRGYVLEPGES
LVTEGRKNEREVAARFASPDADAYASNSAAGDSRLNGVIGVALFELDAPASGREAPA
DGPQAFPADARNGGYPAPPRYRD"
887. .1387
/gene="PA3318"
887. .1387
/gene="PA3318"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06706.1"
/db_xref="GI:9949448"
/translation="MNRFSUPRDPARQCVAGLGSFALLLLASSANAAACRDTITSVQOQS
ADARTNSMGHVTQHIYGTTPSSQSKDTLFSKRGNDYAAARQYQYIDNPVACSG
NSKFEQVSLKHLHMKIDAYSCQADQGGACTRWDTYMAKETSIGFVLKDGKWLITL
NVPVPLD"
/gene="plcN"
complement(1445. .3523)
/gene="plcN"
/notes="PA3319"
complement(1445. .3523)
/gene="plcN"
/codon_start=1
/transl_table=11
/product="non-hemolytic phospholipase C precursor"
/protein_id="AAG06707.1"
/db_xref="GI:9949449"
/translation="MISKRRSFIRLAAGTATGATVATSMPLSSIOAALAIHAHRHCHN
LKDEHVILMOENRSDHYFTLKGVCGRDMAIPLDGOVWVHOKSGKEGFLPYH
FDISTTSAQVDGTPHTPDQAQWNEGRMDKWLPAKTERSLGYKEQDIAQFAMAN
AFICIDAYCSFGQGTNPRLFTGTNDPLQGHGGVPTNDHDSNGPVEQGYTWYTY
PERLQAQITWRYQDMNDFSDNPLIGFRQYRAAADPSPLVNGLSTWKLDAIKRDY
LANSLPQSVIVAPAKYSEHPGSPPIGAETSWLDALTANPEVWSKTALLVMFDE
NDGFFDHVAPPAAPSLANKDGLRGKTADATLEWHTKGDIRYNQOPYGLGPRVPMYVI
SPMSKGGWNSQVFDHTSVIRFLQRFQVMEPNISPMWRRAVCCDLTSAPFNANPNRP
FPELPDTQADAIVASQIKLPKPKPAVAMPKQEMGIRPARALPYELGVHARYRSGG
DALSLFTANTGKAGAVQFVDLLDSENPKRYTVGARKRLHDSFOGDASGDYHLEVIG
PNGFLRVFGRNLRDLAERKAPLPEVRIIDYEPLEGNLRVQLINRGRHPVKLTVKDNVY
RQERTYVNPQGRORREVRISLSSGNWYDFSAQADSF LRRFSGRMEDGRSGFSD
PGMGLGTLFF"
complement(3754. .4209)
/gene="PA3320"
complement(3754. .4209)
/gene="PA3320"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06708.1"
/db_xref="GI:9949450"
/translation="MPNYLMVMAVSGVMVVLQAGSNALLGRLLGHPLSLASLTLGCV
SLGVVSLGLLVRLPPLATASLGGAPWMSWLGGLFGAAYLTVALVAPRLGAATFIAC
VWAGOMLVSAICDQFGWAGFPVRRLSWEGYGLGLVALVVGAVLLLOWRGA"
4298. .5212
/gene="PA3321"
4298. .5212
/gene="PA3321"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"

```

BASE COUNT 1686 a 3490 c 3309 g 1607 t

## ORIGIN

## alignment\_scores:

Quality: 271.50 Length: 462  
 Ratio: 1.201 Gaps: 20  
 Percent Similarity: 48.918 Percent Identity: 25.541

## alignment\_block:

US-09-426-072-2 x AE004754/rev ..

Align seg 1/1 to reverse of: AE004754 from: 1 to: 10092

46 AsnLeuLysAspLysValGluAsnIleValTrpLeuIleLeuGluAsnAr 62  
 3394 ACCCTCAAGGAC...GTCCAGCAGCTGTGCTCTGATCGAGGAGACCG 3348  
 62 gSerPheAspAsnIleLeuGly.....GlyValArgArgGlnGly. 75  
 3347 CTCCTTCGACCACTATTTCGGCACCCTCAAGGGCGTCCGGCTTCGGCG 3298  
 76 .....LeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrIysAsn 90  
 3297 ACCGATGGCGATCCCGCTGCCGATGGCCAGCGGTCTGGCACCAGAAAG 3248  
 91 AlaSerAspProSerSerGlyLysTyr.....CysThrG1 102  
 3247 GGCAGCAAGGCGGAGATCTGTCCTTACCCTTCGACACAGCACCACCAAG 3198  
 102 nAlaLysAspTyrAspSerValPheAsnAspProAspHisSerValThrG 119  
 3197 CGCCAGCGGCTCGACGCG..... 3179  
 119 LysAsnAsnLeuGluPheTyrGlyThrTyrProAsnAsnGlyAlaIle 135  
 3178 .....ACCCGCGACACCTGGCGCGGAC 3158  
 136 AlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLe 152  
 3157 GCC.....CAGCAGCGCTTGGAAACGAAGGCGCAT 3129  
 152 uAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnValMetG 169  
 3128 GGAC.....AAGTGGCTCGCGGCCAAG...ACCGAGCGTTCCTCTGG 3091  
 169 LysTyrThrGluGluGluValProThrLeuValAspLeuValAspGlu 185  
 3090 GCTACTACAGGACGAGGATCGCTTCAGTTCGCGATGCCACACGCC 3041  
 186 PheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsnPr 202  
 3040 TTCACCATCTCGCAGCGCTATCATCTGCTTCGAGGGCGGCGACCAACC 2991  
 202 oAsnArgLeuCysAlaLeuAlaGlyThr.....AlaAlaGlyHisG 216  
 2990 CACACGGCTGTCTCTGAGCGGCGACCAACGACCGCTCGGCGGACGACG 2941  
 216 Lys.....LysAsnAspAspPhe.....LeuAsnTyr 225  
 2940 GTGGTCCGGTAACCAACAGCAGCAGACAGCAGCGCGCGGTGGAGCAG 2891  
 226 GlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLysGlyValSe 242  
 2890 GCGTACACTGGACCACTATCCGAGCGGCTCGAGGCTGGCGGCGGATCAC 2841  
 242 rTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAspSerLeu. 258  
 2840 CTGGGGGTCTTACCGAGACATGGCGGACAACTTCTCCGACAAACCGGTGA 2791  
 259 .....PhePheThrTyrValAsnGlnThrSerArgSerAsnValVal... 272  
 2790 TCGGCTTCGCGCAGTACCGCGCGGCGGCTGACTCGCGCTGATCGTC 2741

273 .....ProValGluAsnPhePheGlnAspAlaTyrLe 283  
 2740 AACGGCTGAGCACCTGGAAAGCTCGATCGCTGAAGCGGACGCTGCTGGC 2691  
 283 uClyValLeuProLysPheSerTyrIleAsnProSerCysGlyThrA 300  
 2690 CATAGCTGCGCGCAGGTGCTCTGTGATC.....GTCCGCGCGG 2653  
 300 snThrAsnSerMetHisPro...ThrGlyAsnValSerTyrGlyGluVal 315  
 2652 CCAAGTACTCTCCGAACACCGCGCGCTCCAGCCGATCTGGGGGCGCGAG 2603  
 316 PheValLysGlnIleTyrAspAlaIleArgGlnGlyProGln...TrpAs 331  
 2602 TACACTTCCCTGGGTACTCGACGCGCTACCGCCCAACCCGGAGGCTCGAG 2553  
 331 pLysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheTyrAspH 348  
 2552 CAAGACCGCGCTGCTGGTGATGTTGACGAGAAGACGAGCTTCTTCGACC 2503  
 348 IsValProProProLeuAla.....ValArgProAspAsnLeuThrTyr 362  
 2502 ACGTCGCGCGCGCGCGCGCTGAAACAAGGACGCGCACGCTCGCT 2453  
 363 ThrGluThr.....AlaLysAsnGlyG1 370  
 2452 GCAAGACACCGCGCGCGCGCTCGAATGGCACACCAAGGCGGATAT 2403  
 370 nLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTrpV 387  
 2402 CGTTATCGCAACCAACCGCTAGCGCTCGCGCGCGGTGCGGATGACG 2353  
 387 alIleSerProTyrSerLysGlyTyrIleGluGlnTyrGlyThrAsp 403  
 2352 TGATCTCGCGTGGAGCAAGGCGCTGGGTCAACTCCCA.GGTGTTCCA 2304  
 404 ProValThrGlyLysProAlaProTyrSerAlaThrSerValLeuLysTh 420  
 2303 CCACACCTCGGTATCCGCTTCCTGGAGCAGCGCTTCGG..... 2265  
 420 rLeuGlyTyrLeuTrpAspIleGluAspPheThrProArgValAlaHis 437  
 2264 .....GGTCATG 2258  
 437 erProSerPheAspHisLeuIleGlyThrThrLeuArgGlu...AspAla 452  
 2257 GAGCGGAATATCAGTCCCTGGCGCTGCTCGCGTCTCGGCGGACCTGACCTC 2208  
 453 ProfileAlaLeuLysThrProHisThrPheSerVal 464  
 2207 GCGCTTCAACTTCGCCAACCGCAACACGAGCGGTT 2172

OM of: US-09-426-072-2\_COPY\_20\_464 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Feb 27, 2001 12:47 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame1.p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/6146869/runat_13022001_095136_15010/app_query.fasta_1.1049
-DB=N_Geneseq_36 -QFWT=fastap -SUFFIX=ring -CAPOP=12,000
-GAPEXT=4,000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-XGAPEXT=4,500 -XGAPEXT=0.050 -XGAPEXT=10.000 -XGAPEXT=0.500
-CAPOP=6,000 -FGAPEXT=7,000 -YGAPOP=10,000 -YGAPEXT=0.500
-DELOP=6,000 -DELEXT=7,000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=6146869 -CGN1_1_170 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-426-072-2_COPY_20_464
Query length: 445
Database: N_Geneseq_36:*
Database sequences: 480022
Database length: 187831343
Search time (sec): 114.620000
```

score\_list:

Sequence	Strd Orig	zscore	EScore Len	Documentation
/cgn2_2/gcgdata/geneseq/NA1990.DAT:Q06266 +	126.50	204.47	0.0013	1770
/cgn2_2/gcgdata/geneseq/NA1997.DAT:T44068 +	122.00	186.45	0.0135	4360
/cgn2_2/gcgdata/geneseq/NA1997.DAT:T61372 +	111.50	168.38	0.1368	3876
/cgn2_2/gcgdata/geneseq/NA1993.DAT:Q42754 +	109.50	171.51	0.0916	2062
/cgn2_2/gcgdata/geneseq/NA1999.DAT:X60600 +	109.50	167.02	0.1629	3126
/cgn2_2/gcgdata/geneseq/NA1999.DAT:X60602 +	109.50	163.59	0.2528	4292
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X49213 +	107.00	165.90	0.1881	2264
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X49214 +	107.00	165.90	0.1881	2264
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X49212 +	107.00	165.69	0.1931	2307
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X49209 +	107.00	158.12	0.5101	4652
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X34958 +	106.50	167.96	0.1452	1718
/cgn2_2/gcgdata/geneseq/NA1999.DAT:Z07051 +	105.50	159.65	0.4192	3126
/cgn2_2/gcgdata/geneseq/NA1999.DAT:Z30163 +	105.00	132.92	12.92	34094
/cgn2_2/gcgdata/geneseq/NA1995.DAT:Q91398 +	104.50	157.92	0.5234	3094
/cgn2_2/gcgdata/geneseq/NA1995.DAT:Q83484 +	104.50	157.69	0.5387	3159
/cgn2_2/gcgdata/geneseq/NA1997.DAT:T93873 +	104.50	157.69	0.5387	3159
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X60181 +	104.50	157.69	0.5387	3159
/cgn2_2/gcgdata/geneseq/NA1995.DAT:Q83489 +	104.50	148.70	1.71	7266
/cgn2_2/gcgdata/geneseq/NA1997.DAT:T93872 +	104.50	148.70	1.71	7266
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X260182 +	104.50	148.70	1.71	7266
/cgn2_2/gcgdata/geneseq/NA1996.DAT:T30656 +	104.50	146.83	2.17	8640
/cgn2_2/gcgdata/geneseq/NA1998.DAT:X54296 +	104.00	159.96	0.4029	2352
/cgn2_2/gcgdata/geneseq/NA1993.DAT:Q42760 +	103.50	161.81	0.3179	1820
/cgn2_2/gcgdata/geneseq/NA1987.DAT:N70818 +	103.50	160.46	0.3779	2062
/cgn2_2/gcgdata/geneseq/NA1999.DAT:X60001 +	103.50	158.02	0.5168	2062
/cgn2_2/gcgdata/geneseq/NA1984.DAT:N40068 +	103.00	159.55	0.4247	2585
/cgn2_2/gcgdata/geneseq/NA1992.DAT:Q27516 +	102.00	162.69	0.2840	2060
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X54295 +	102.00	156.27	0.6462	1299
/cgn2_2/gcgdata/geneseq/NA1996.DAT:T16803 +	101.50	155.88	0.6801	2352
/cgn2_2/gcgdata/geneseq/NA2000.DAT:X236889 +	101.50	153.84	0.8827	2241
/cgn2_2/gcgdata/geneseq/NA1998.DAT:X14547 +	101.00	140.25	5.05	2705
/cgn2_2/gcgdata/geneseq/NA1998.DAT:X52091 +	101.00	140.25	5.05	8748
/cgn2_2/gcgdata/geneseq/NA1992.DAT:Q31411 +	100.50	148.41	1.77	9748
/cgn2_2/gcgdata/geneseq/NA1992.DAT:Q32161 +	100.50	148.41	1.77	3771
/cgn2_2/gcgdata/geneseq/NA1992.DAT:Q30937 +	100.50	148.41	1.77	3771
/cgn2_2/gcgdata/geneseq/NA1993.DAT:Q51689 +	100.50	148.41	1.77	3771
/cgn2_2/gcgdata/geneseq/NA1995.DAT:Q94055 +	100.50	148.41	1.77	3771
/cgn2_2/gcgdata/geneseq/NA1997.DAT:T60071 +	100.50	148.41	1.77	3771

```
/cgn2_2/gcgdata/geneseq/NA1990.DAT:Q01811 + 99.50 151.51 1.19 2388
/cgn2_2/gcgdata/geneseq/NA1992.DAT:Q20336 + 99.50 146.57 2.24 3771
/cgn2_2/gcgdata/geneseq/NA1997.DAT:Y25144 + 99.00 146.03 2.41 3642
/cgn2_2/gcgdata/geneseq/NA1992.DAT:Q30805 + 98.50 144.73 2.84 3771
/cgn2_2/gcgdata/geneseq/NA1999.DAT:X13377 - 98.50 133.57 11.88 10594
```

seq\_name: /cgn2\_2/gcgdata/geneseq/NA1990.DAT:Q06266

seq\_documentation\_block:

```
ID Q06266 standard; DNA; 1770 BP.
XX
AC Q06266;
XX
DT 01-FEB-1991 (first entry)
XX
DE Promoter region and sequence encoding secretory signal peptide.
XX
KW Acid phosphatase; yeast; ds.
XX
OS Saccharomyces cerevisiae.
XX
FH Key Location/Qualifiers
FT Promoter 1..132
FT CDS /tag= a
FT 133..195
FT /tag= b
XX
PN JP02234681-A.
XX
PD 17-SEP-1990.
XX
PF 09-MAR-1989; 89JP-0057234.
XX
PR 09-MAR-1989; 89JP-0057234.
XX
PA (IKEM ) IKEDA MOHANDO KK.
XX
DR WPI; 1990-325618/43.
XX
DR P-PSDB; R07445.
XX
PT DNA contg. base sequences with promoter activity - used for
expression and secretory prodn. of pharmacologically important
protein gene
XX
PS Claim 1; Page 553; 11pp; Japanese.
XX
CC Sequence has promoter activity and encodes a secretion signal
peptide. Heterologous constructs may be produced, expressible in
yeast, as a high yield secretory product.
XX
SQ Sequence 1770 BP; 507 A; 382 C; 324 G; 557 T; 0 other;

alignment_scores:
Quality: 126.50 Length: 291
Ratio: 0.861 Gaps: 14
Percent Similarity: 50.515 Percent Identity: 23.024

alignment_block:
US-09-426-072-2_COPY_20_464 x Q06266 ..
Align seg 1/1 to: Q06266 from: 1 to: 1770

115 ALaLeaSerGlyLysValValAlaAspGln...SerGlyPheLeuAs 130
|||||...|||||...
274 GCTGCTGCTTCTCTCTGCTGCTCAGCAAGAACTCGGACGTAAGGG 323
130 nAlaGlnLeuAsn.....AspTyrP 137
|||||...|||||
324 TCGGCGTGTCAACCGTTTTCAGTCATCTTCTTGGAAACACAGATTATG 373
137 rOLysLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrThr 153
|||...:|||||
```





```

86 AsPTyRAsSerValPheAsnAspProAspHisSerValThrGlyAsnAs 102
   ::::: ::::: ||||| ||| :::::
2199 AATCATCTCGGAGGCGTGTACCCCAATAGGAGTGAATTTCCCAAGA 2248
   ::::: ::::: ||||| ||| :::::
102 nLeuGluPheTyRThrProAsnAsnGlyAlaIleAlaSerG 119
   ::::: ||| ||||| :::::
2249 TATTT.....TATATCCCTAGA.....TCCTCAT 2271
   ::::: ::::: ||||| ||| :::::
119 lyLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAsp 135
   ::::: ::::: ||||| ||| :::::
2272 CTAACATCATCTTGGCAATATCTGCT.....ATGAATAGC 2306
   ::::: ::::: ||||| ||| :::::
136 TyrProLysLeuAlaProGluAlaThrArgGlnVal.....MetG1 150
   ::::: ::::: ||||| ||| :::::
2307 TCAGTGGAGATTAAACAGCGGATGGACTTTAATGTTTACCACCTGG 2356
   ::::: ::::: ||||| ||| :::::
150 yTyRThrGluGluGluValProThrLeuValAspLeuValAspGluP 167
   |||| ::::: ||||| ||| :::::
2357 ATATATCAATGATGAATCCATA.....TTGAACAATGCAATTAAGATC 2400
   ::::: ::::: ||||| ||| :::::
167 heThrThrPheAsnSerTrpPheSerCysVal.....ProGlyProThrAsn 182
   ::::: ::::: ||||| ||| :::::
2401 TTTATTTATTTTCAGAAAGACAGCTCGCTGCGCTGCAATGACATAT 2450
   ::::: ::::: ||||| ||| :::::
183 ProAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGly..... 197
   |||| ::::: ||||| ||| :::::
2451 CCGGTTTTTGTGGCGAGCTGCTGCT...CTGCTGAATCAGCGCTTTC 2497
   ::::: ::::: ||||| ||| :::::
197 ..... 197
2498 TTCTCAAGTTACTTATATAATCGGTTCTTCCATCTTACGGGCTCCTT 2547
   ::::: ::::: ||||| ||| :::::
198 .....LysAsnAspAspPheLeuAsnTyrGlyIle 208
   ::::: ::::: ||||| ||| :::::
2548 TCGCTCAATTTAGTGAGCAATCAGACGACACTTTTAAACAACGGATTA 2597
   ::::: ::::: ||||| ||| :::::
209 SerSerLysSerIle...PheGluAlaAlaAsnGluLys..... 220
   ::::: ::::: ||||| ||| :::::
2598 ACCAGCCAGCATTCCTTTTAAACAGCTAACGGTGATTTCTACAGAG 2647
   ::::: ::::: ||||| ||| :::::
221 .....GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluP 234
   |||| ::::: ||||| ||| :::::
2648 CATTCTGTTTGGGTTTAAACGAATCCGATAT.....TCTTATGAGG 2688
   ::::: ::::: ||||| ||| :::::
234 heGluProAspSerLeuPheThrTyrValAsnGlnThrSerArgSer 250
   ::::: ::::: ||||| ||| :::::
2689 TTGATCCAGATCT.....AAAAAATTACCGTTTGTAAAGTTC 2729
   |||| ::::: ||||| ||| :::::
251 AsnValValProValGluAsnPheGlnAspAlaTyrLeuGlyValLe 267
   |||| ::::: ||||| ||| :::::
2730 AAT.....CCAATAGAA.....CTACCGTTGCT 2752
   ::::: ::::: ||||| ||| :::::
267 uPro.....LysPheSerTyrIleAsnPro..... 275
   |||| ::::: ||||| ||| :::::
2753 CCCTGGTGGTATCGCTATTAGAAACTTCAATATATGACCCAGTTTATAG 2802
   ::::: ::::: ||||| ||| :::::
276 .....SerCysCysGlyThrAsnThrAsnSerMetHisProThrGlyAsn 290
   |||| ::::: ||||| ||| :::::
2803 ATATAATAATTGACGACCAACATGGTACGATTTGTTCATAATCAGGAGAT 2852
   ::::: ::::: ||||| ||| :::::
291 ValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgG1 307
   |||| ::::: ||||| ||| :::::
2853 GTT.....CCTATTATATAAAG..... 2870
   ::::: ::::: ||||| ||| :::::
307 nGlyProGlnTrpAspLysThrLeuPheIleThrTyrAspGluThrG 324
   ::::: ::::: ||||| ||| :::::
2870 ..... 2870
324 lyGlyPheTyrAspHisValProProLeuAlaValArgProAspAsn 340
   ::::: ::::: ||||| ||| :::::
2871 .....ATACCAACAGATCTCTAATACATGACCAAGGAT 2903
   ::::: ::::: ||||| ||| :::::
341 LeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHis... 356

```

```

::: ::::: ||| ::::: ||||| ::::: |||||
2904 ATCAACTTCTATAATGTTCCGAAAACGAAAGAAAACCAATCTAGAGCG 2953
   ::::: ::::: ||||| ||| :::::
357 .....PheAspArgLeuGlyGlyArgMetProThrTrpValIleSerP 371
   ||||| ::::: ||||| |||
2954 TAGAGACGTCGACCGTGTGGTGATCCAATG..... 2984
   ::::: ::::: ||||| |||
371 roTyRSerLysLysGlyTyrIleGluGlnTyrGlyThr 383
   ::::: ::::: ||||| |||
2985 .....AGGATGGATAGGTATGGTACC 3005

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/NA1993.DAT:Q42754

seq\_documentation\_block:

ID Q42754 standard; DNA; 2062 BP.

XX Q42754;

XX DT 25-NOV-1993 (first entry)

XX pSY875 swine parvovirus B gene sequence.

XX Attenuated; vaccine; herpes virus; non-primate; live; safer; IBR;

XX Infectious bovine rhinotracheitis; MDV; Marek's disease virus; fowl;

XX Swine parvovirus.

XX Key Location/Qualifiers

XX CDS 267..2006

XX FT /\*tag= a

XX US5223424-A.

XX PD 29-JUN-1993.

XX PF 27-JUL-1988; 88US-0225032.

XX PR 06-SEP-1985; 85US-0773430.

XX PR 27-JAN-1986; 86US-0823102.

XX PR 17-JUL-1986; 86US-0887140.

XX PR 02-SEP-1986; 86US-0902887.

XX PR 20-NOV-1986; 86US-0933107.

XX PR 27-JUL-1987; 87US-0078519.

XX PR 27-JUL-1988; 88US-0225032.

XX PA (PRUT-) PRUTECH RES & DEV.

XX PI Chiang CH, Cochran MD, Macdonald RD;

XX DR WPI; 1993-219585/27.

XX DR P-PSDB; R38697.

XX PT Recombinant fusion proteins for vaccine - comprises antigenic

XX PT sequences fused to viral sequences e.g. pseudo-rabies virus, used

XX PT as vaccines

XX PT Disclosure; Fig 11; 127pp; English.

XX PS The sequence is that of the swine parvovirus B gene sequence from

XX CC pSY875 which may be used in the prodn. of attenuated non-primate

XX CC herpes viruses. These can be used as live vaccines and provide a

XX CC safer vaccine than currently available for e.g. pseudorabies virus

XX CC of swine, infectious bovine rhinotracheitis (IBR) virus or Marek's

XX CC disease of fowl.

XX SQ Sequence 2062 BP; 818 A; 447 C; 334 G; 463 T; 0 other;

alignment\_scores:

Quality: 109.50

Length: 402

Ratio: 0.622

Gaps: 22

Percent Similarity: 43.781

Percent Identity: 21.393

```

alignment_block:
US-09-426-072-2_copy_20_464 x Q42754 ..
Align seg 1/1 to: Q42754 from: 1 to: 2062

91 PheAsnAspProAspHisSerValThrGlyAsnAsnLeu..... 103
||||| ||| ..... ||||| |||
627 TTCATCCGCGGACTGGCAGTTATATATCAACAACATGACAGAAATAA 676
104 .....GluPheTyr..... 106
677 CTTAGTTAGTTTTGAACAAGAAATATTCATGTAGTACTTAAACAATTA 726
107 .....GlyThrTyrThrPro.....AsnAsnGlyAlaIle 116
||||| ||| |||||
727 CAGATACAGCACTCACCACCATCCCAATATATATGATCTAACT 776
117 AlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLe 133
||||| ||| ..... |||
777 GCAAGCTTAATGTGTCGCTAGACACCAATAACACACTT..... 815
133 uAsnAspTyrProLysLeuAlaProGluAlaThrArgGlnValMetG 150
||| ||||| ..... |||
816 .....CCATACACACCAGCACCTAGAGCTGAACA.....CTTG 852
150 lyTyrTyrThrGluGluGluValProThrLeuValAspLeuValAspGlu 166
||||| ||| ..... |||
853 GTTTTAT.....CCATGTTTACTACA.....AAA 878
167 PheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsnPr 183
||| ..... ||||| ..... |||
879 CCAACTCAATACAGATATTACCTATCATGCTCAGAAACCTAAATCCACC 928
183 oAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsnA 200
| ..... ||| ..... |||
929 A.....ACATACACTGGACAATCAACAACAA 954
200 sPAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePheGluAla 216
||| ..... ||| ..... |||
955 TAACAGACTCAATACAAACAGGACTACACAGTGACATTATGTCTACACA 1004
217 AlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGly.....Th 230
..... |||
1005 ATAGAAAATGCGTACCAATTCATCTTCTAAGAACTGGAGATGATTC 1054
230 rAsnGly.....GluPheGluProAspSerLeuPhePheThrTyrValA 245
||||| ..... ||| ..... |||
1055 CACAGGAATATATCACTTTGACACAAACCAATTAATACTCCTCAT 1104
245 snGlnThrSerArgSerAsnValProValGluAsnPhePheGln.... 260
||||| ..... ||| ..... |||
1105 GGCAACAACAGATCTCTAGAGCTGCTCCANAGTACTACTAAGTAACT 1154
261 .....AspAlaTyrLeuGlyValLeuProLysPheSerTyrIle 273
||| ||| ..... |||
1155 ACCACAGAAGGAGACCAACACCCAGGACACTACCA..... 1190
273 eAsnProSerCysCysGlyThrAsnThr...AsnSerMetHisProThrG 289
||| ..... |||
1191 .....GGAGCTACACACAGAAAGGTTATACCAACAA 1224
289 lysAsnValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIle 305
||| ||||| ..... |||
1225 TTAATAATAGCTACACAGAAGCA.....ACAGCACTT 1256
306 ArgGlnGlyPro...GlnTrpAspLysThrLeuPheIleThrTyrAs 321
||| ..... |||
1257 AGCCAGCTAGGTAGATATATACACCATACATGATTTGACTACTC 1306
321 pGluThrGlyGlyPheTyrAspHisValProProProLeuAlaValArgp 338
||| ||| ..... |||
1307 CAATGGTGGACCACTTCTA.....ACTCCTATAGTACCAACAG 1344

```

```

338 roAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThr 354
||||| ||| ..... ||||| |||
1345 CAGACACACAATATATGATGATGAACCAATGTGCTATAGAATTACA 1394
355 Leu.....HisPheAspArgLeuGlyGlyArgMetProTh 366
||| ..... ||| ..... |||
1395 ATGGGTTACACACATGACACCTTAACACACATCTTCAAGAGCTAGAAAG 1444
366 rTrpValIleSerProTyrSerLysLysGlyTyrIle.....GluGlnT 381
||||| ..... ||| ..... |||
1445 ATACACATTAATCCACAAGTAATGTGAAGAGCTCCAAAGCAACAAT 1494
381 yFGly..... 382
1495 TTAATACACAGGCACCACTAAACCTAGNAAATACAAATTAATGGAACACTT 1544
383 .....ThrAspProValThrGlyLysPro..... 390
||||| ..... |||||
1545 TTACCTTCAGATCCAATAGGAGGAAATCTAACAAGCATTTTCATGAATAC 1594
391 .....AlaProTyrSerAla.....ThrSerValLeuL 400
||| ..... ||| ..... |||
1595 ACTCAATACATATGGACCATTAACAGCACATAACAATACTGCACCTGTAT 1644
400 yThrLeuGlyTyrLeuTrpAspIleGlu.....AspPheThrPro 413
||| ..... ||| ..... |||
1645 TTCCAATGTGTCAAATATGGGATAAAGAACTTGATACAGATCTAAACCT 1694
414 ArgVal 415
|||||
1695 AGACTA 1700

seq_name: /cgr2_2/cgdata/geneseq/geneseq/NAL1999.DAT.X60600
seq_documentation_block:
ID X60600 standard; DNA; 3126 BP.
XX
AC X60600;
XX
XX 27-JUL-1999 (first entry)
XX
DE S. cerevisiae acidic trehalase (ATH1) gene sequence.
XX
KW Acidic trehalase; ATH1 gene; gene manipulation; baker's yeast; dough;
KW fermentation; baking; freezing; bread; trehalose degradation; ss.
XX
OS Saccharomyces cerevisiae.
XX
PN EP921190-A2.
XX
PD 09-JUN-1999.
XX
PF 04-DEC-1998; 98EP-0403046.
XX
PR 08-DEC-1997; 97JP-0352016.
XX
PA (NORQ ) NAT FOOD RES INST MIN AGRIC.
PA (ORIV ) ORIENTAL YEAST CO LTD.
XX
PI Iyo C, Mori K, Nakajima R, Shima J, Suzuki Y, Takano H;
PI Watanabe H;
XX
XX WPI; 1999-315405/27.
XX
XX P-PSDB; Y16788.
XX
XX New acid trehalase (ATH1) gene-disrupted haploid yeast, useful for
XX preparing high-sugar dough resistant to freezing
XX
XX Examples; Page 16-23; 57pp; English.
XX
XX The invention relates to an acid trehalase (ATH1) gene-disrupted haploid
XX yeast produced by gene manipulation, where the diploid is a practical
CC

```







```

|||||
3424 TTGATAATCAGGAGATGTTCCATTATCATATAAAGATACCAACACAGATCT 3473
|||||
335 AlavalatgProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyG1 351
|||||
3474 CTAATACATGACCGAGGATATCAACTTCTATATAGTTCGGAAGAAGAAG 3523
|||||
351 nLysTyrThrLeuHis.....PheAspArgLeuGlyGlyArgMetP 365
|||||
3524 AAAACCAATCTAGAGGCTAGAGACGTCGACCGTGTGGTGTATCCAATG. 3572
|||||
365 rothrtrpValIleSerProTyrSerLysLysGlyTyrIleGluGlnTyr 381
|||||
3573 .....AGGATGGATAGGTAT 3587
|||||
382 GlyThr 383
|||||
3588 GGTACC 3593

```

seq\_name: /cgn\_2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:Z49213

seq\_documentation\_block:

ID: Z49213 standard; DNA; 2264 BP.

XX AC Z49213;

XX XX 15-MAR-2000 (first entry)

XX XX Adeno associated virus AAV5 capsid protein VP2 DNA.

XX XX Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;  
 KW promoter; Rep protein; capsid protein; regulation; transcription;  
 KW replication; chromosomal integration; tissue tropism; cellular receptor;  
 KW gene therapy; neutralising antibody; erythroid progenitor cell;  
 KW transduction; cancer; genetic disease; VP2; ss.

XX OS Adeno associated virus type 5.

XX XX Key Location/Qualifiers

FT CDS 470..2236

FT FT /\*tag= a

FT FT /product= "AAV5 capsid protein VP2"

FT FT /note= "The translation initiation codon in this  
 sequence is ACG, which encodes Thr; this  
 unusual start codon is specifically mentioned in  
 the specification"

XX FT WO9961601-A2.

XX XX 02-DEC-1999.

XX PF 28-MAY-1999; 99WO-US11958.

XX PR 28-MAY-1998; 98US-0087029.

XX XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX XX Chiorini JA, Kotin RM;

XX XX WPI; 2000-062707/05.

XX DR P-PSDB; Y58161.

XX PT Adeno-associated virus 5 based vectors and particles, useful for gene  
 therapy -

XX PS Claim 31; Page 83; 91pp; English.

XX XX This sequence represents DNA encoding the VP2 capsid protein of  
 CC adeno associated virus type 5 (AAV5). The invention relates  
 CC to vectors comprising a pair of AAV5 inverted terminal repeats  
 CC (ITRs) with a promoter between the ITRs. The vector may comprise  
 CC the viral genome, or subregions thereof, including sequences  
 CC encoding Rep proteins and capsid proteins, and is encapsidated

CC in an AAV5 particle. The non-structural Rep proteins Rep40  
 CC (Y58613), Rep52 (Y58168), Rep68 (Y58164) and Rep78 (Y58159) are  
 CC involved in regulation of replication and transcription, in addition to  
 CC the production of progeny genomes. Rep68 and Rep78 are also associated  
 CC with the stable integration of the viral genome into human chromosomes.  
 CC The three types of capsid protein VP1 (Y58160), VP2 (Y58161) and VP3  
 CC (Y58162) assemble to form an icosahedral capsid, and differ from each  
 CC other by the use of alternative splicing and an unusual translation  
 CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2  
 CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are  
 CC likely to utilise distinct cellular receptors and are serologically  
 CC distinct. In a gene therapy application, therefore, AAV5 would allow for  
 CC transduction of a patient who already possess neutralising antibodies  
 CC either as a result of natural immunological defence or from prior  
 CC exposure to AAV2 vectors. The vectors may be useful for transducing  
 CC erythroid progenitor cells or cells lacking heparin sulphate  
 CC proteoglycans, which is very inefficient with AAV2-based vectors.  
 CC The vectors may also be useful for transducing cells with a nucleic  
 CC acid of interest in order to produce cell lines that could be used to  
 CC screen for agents that interact with the gene product of the nucleic  
 CC acid of interest. In addition to transduction of other cell types,  
 CC transduction of erythroid cells would be useful or the treatment of  
 CC cancer and genetic diseases which can be corrected by bone marrow  
 CC transplants using matched donors.

XX XX

SO Sequence 2264 BP; 575 A; 722 C; 595 G; 372 T; 0 other;

alignment\_scores:

Quality: 107.00 Length: 346  
 Ratio: 0.738 Gaps: 16  
 Percent Similarity: 41.908 Percent Identity: 19.075

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x Z49213 ..

Align seg 1/1 to: Z49213 from: 1 to: 2264

```

99 ThrGlyAsnAsnLeuGlu.....PheTy 106
|||||
1253 ACGGGACAACCTTTGAGTTTACCTACAACTTTGAGGAGTGCCCTTCCA 1302
|||||
106 rGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValValA 123
|||||
1303 CTCAGCTCGCTCCAGTCAGAACTGTTCAAGCTGGCAACCGCTGG 1352
|||||
123 laAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeu 139
|||||
1353 TGGACCACTACTTGTACCGCTTCGTGAGCACAATAAC.....ACT 1393
|||||
140 AlaProGluGluAlaThrArgGlnValMetGlyTyrThrGluGluG1 156
|||||
1394 GCGGAGTCCAGTTCAACAGAACCTGGCGGAGATACGCC..... 1435
|||||
156 uValProThrLeuValAspLeuValAspGluPheThrThrPheAsnSert 173
|||||
1436' .....AACCTCAAAACT 1451
|||||
173 rpPheSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeu 189
|||||
1452 GGTTTC.....CCGGGCCCATGGCGCAACCCAG..... 1480
|||||
190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206
|||||
1481 ...GGCTGGACCTCGGCTCCGGGTCAACCGC..... 1510
|||||
206 rGlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLysGlyValS 223
|||||
1511 .....GCCAGTGTACAGCGCTTCGCCACGACCAATAGG..... 1543
|||||
223 erTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAspSerLeu 239
|||||
1544 .....ATGGAGCTCGAGGCGCGAGTACCAGGTGCCCGCGCAGCGAAC 1588

```

```

240 PhePheThrTyrValAlaGlnThrSerArgSerAsnValProValGI 256
XX      :|||||  |||:  :|:  |||||:  :|:  |||
1589 GCGATGACC.....AACAACTCCAGGCGAGCAACACCTATGCCCTGGA 1632
XX      :|||||  |||:  :|:  |||||:  :|:  |||
256 uAsnPhePhe.....GlnAspAlaTyrLeuGlyValLeuProL 269
XX      :|:  :|:  |||  |||  |||  |||  |||  |||
1633 GAACACTATGATCTTCAACAGCAGCGCGGCAACCGGCGCACCGCC. 1681
XX      :|||||  |||:  :|:  |||||:  :|:  |||
269 ysPheSerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMet 285
XX      :|||||  |||:  :|:  |||||:  :|:  |||
1682 .....ACCTACCTCGAGGCGCAACATGCTCATCCAGCAGGAGCAGC 1726
XX      :|||||  |||:  :|:  |||||:  :|:  |||
286 HisProThrGlyAsnValSerTyr..... 293
XX      :|||||  |||:  :|:  |||||:  :|:  |||
1727 CAGCGGTGAACCGCTGCGGTACAACTCGGCGGCGAGATGCCACCAA 1776
XX      :|||||  |||:  :|:  |||||:  :|:  |||
293 ..... 293
XX
1777 CAACAGAGCTCCACCACTGCCCGCGGCGGCGGCGGCGGCGGCGG 1826
XX
294 .....GlyGluValPheVallyGlnIleTyrAspAlaIleArg 306
XX      :|||||  |||:  :|:  |||||:  :|:  |||
1827 AAATCGTCCCGCGCAGCGTGTGGATGGAGAGG.....GACGTGTACCTC 1870
XX      :|||||  |||:  :|:  |||||:  :|:  |||
307 GlnGlyProGlnTrpAspLysThrLeuPheIleThrTyrAspGluTh 323
XX      :|||||  |||:  :|:  |||||:  :|:  |||
1871 CANGAGCCCATCTGGGCCAAG.....ATCCAGAGAGC 1902
XX      :|||||  |||:  :|:  |||||:  :|:  |||
323 rGlyGlyPheTyr.....AspH 329
XX      :|||||  |||:  :|:  |||||:  :|:  |||
1903 GCGGCGGCGCACTTTCACCCCTCTCGGCGCATGGCGGATTCGAGCTCA 1952
XX      :|||||  |||:  :|:  |||||:  :|:  |||
329 isValProProLeuAlaValarg.....ProAspAsnLeu 341
XX      :|||||  |||:  :|:  |||||:  :|:  |||
1953 ACCACCGCCCATGATGCTCATCAAGAACAGCGCTGTGCCCGGAATATC 2002
XX      :|||||  |||:  :|:  |||||:  :|:  |||
342 ThrTyrThrGluThrAlaLysAsnGlyClnLysTyrThrLeuHisPheAs 358
XX      :|||||  |||:  :|:  |||||:  :|:  |||
2003 ACC..... 2005
XX
358 pArgLeuGlyGlyArgMetProThrTrpValIleSerProTyrSerLys 375
XX      :|||||  |||:  :|:  |||||:  :|:  |||
2006 .....ACGTTCTCGGAGCGTGCCTGCA 2027
XX      :|||||  |||:  :|:  |||||:  :|:  |||
375 ysGlyTyrIleGluGlnTyrGlyThrAspProValThr 387
XX      :|||||  |||:  :|:  |||||:  :|:  |||
2028 GCAGCTTCATCACCCAGTACAGCACCGGCGGCGAGTCA 2065
XX      :|||||  |||:  :|:  |||||:  :|:  |||

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:249214

seq\_documentation\_block:

ID 249214 standard; DNA; 2264 BP.

XX AC 249214;

XX DT 15-MAR-2000 (first entry)

XX DE Adeno associated virus AAV5 capsid protein VP3 DNA.

XX KW Adeno associated virus type 5.

XX OS 106

XX FH Key Location/Qualifiers

XX FT 638..2236

XX FT /\*tag= a

XX FT /product= "AAV5 capsid protein vp3"

XX FT

XX FT

XX FT

PN W09961601-A2.  
XX  
PD 02-DEC-1999.  
XX  
PF 28-MAY-1999; 99WO-US11958.  
XX  
PR 28-MAY-1998; 98US-0087029.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Chlorini JA, Kotin RM;  
XX  
DR WPI; 2000-062707/05.  
XX  
DR P-PSDB; Y58162.  
XX  
PT Adeno-associated virus 5 based vectors and particles, useful for gene  
PT therapy -  
XX  
PS Claim 32; Page 83-84; 91pp; English.  
XX  
CC This sequence represents DNA encoding the VP3 capsid protein of  
CC adeno associated virus type 5 (AAV5). The invention relates  
CC to vectors comprising a pair of AAV5 inverted terminal repeats  
CC (ITRs) with a promoter between the ITRs. The vector may comprise  
CC the viral genome, or subregions thereof, including sequences  
CC encoding Rep proteins and capsid proteins, and is encapsidated  
CC in an AAV5 particle. The non-structural Rep proteins Rep40  
CC (Y58613), Rep52 (Y58168), Rep68 (Y58164) and Rep78 (Y58159) are  
CC involved in regulation of replication and transcription, in addition to  
CC the production of progeny genomes. Rep68 and Rep78 are also associated  
CC with the stable integration of the viral genome into human chromosomes.  
CC The three types of capsid protein VP1 (Y58160), VP2 (Y58161) and VP3  
CC (Y58162) assemble to form an icosahedral capsid, and differ from each  
CC other by the use of alternative splicing and an unusual translation  
CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2  
CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are  
CC likely to utilise distinct cellular receptors and are serologically  
CC distinct. In a gene therapy application, therefore, AAV5 would allow for  
CC transduction of a patient who already possess neutralising antibodies  
CC either as a result of natural immunological defence or from prior  
CC exposure to AAV2 vectors. The vectors may be useful for transducing  
CC erythroid progenitor cells or cells lacking heparin sulphate  
CC proteoglycans, which is very inefficient with AAV2-based vectors.  
CC The vectors may also be useful for transducing cells with a nucleic  
CC acid of interest in order to produce cell lines that could be used to  
CC screen for agents that interact with the gene product of the nucleic  
CC acid of interest. In addition to transduction of other cell types,  
CC transduction of erythroid cells would be useful or the treatment of  
CC cancer and genetic diseases which can be corrected by bone marrow  
CC transplants using matched donors.  
XX  
SQ Sequence 2264 BP; 575 A; 722 C; 595 G; 372 T; 0 other;

alignment\_scores:

Quality: 107.00 Length: 346  
Ratio: 0.738 Gaps: 16  
Percent Similarity: 41.908 Percent Identity: 19.075

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x 249214 ..

Align seg 1/1 to: 249214 from: 1 to: 2264

```

99 ThrGlyAsnAsnLeuGlu.....PheTy 106
|||||  |||:  :|:  |||||:  :|:  |||
1253 ACGGCGCAACACTTTGAGTTCACCACTTTGAGGAGTGCCTTCCA 1302
XX
106 rGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValValA 123
XX      :|||||  |||:  :|:  |||||:  :|:  |||
1303 CTCGAGCTTCGCTCCCGCAGCAGACCTGTTCAAGCTGGCCACCGCTGG 1352
XX      :|||||  |||:  :|:  |||||:  :|:  |||
123 laAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeu 139
XX

```

```
1353 TGGACCACTTCTTACCGCTCGTGAGCACAATAAC.....ACT 1393
140 AlaProGluGluAlaThrArgGlnValMetGlyTyrThrGluGluG1 156
1394 GCGGAGTCAGTTCACACAGAACCTGGCGGAGATACGCC..... 1435
156 uValProThrLeuValAspLeuValAspGluPheThrThrPheAsnSerT 173
1436 .....AGCTTCTCGGACGTCGCGGAGTCAACCGC..... 1451
173 rpPheSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeu 189
1452 GGTC.....CCGGGCCCATGGCGCAACCCAG..... 1480
190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206
1481 ...GGTGGAACTGGGCTCCGGGCTCAACCGC..... 1510
206 rGlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLysGlyVal 223
1511 .....GCCAGTTCAGCGCTTCGCCACGACCAATAGG..... 1543
223 erTrpLeuAsnTyAspGlyThrAsnGlyGluPheGluProAspSerLeu 239
1544 .....ATGGAGTCGAGGCGCGAGTTACCGAGTCCCGCCGCGACCGCAAC 1588
240 PhePheThrTyrValAsnGlnThrSerArgSerAsnValValProValG1 256
1589 GGCATGACC.....AACAACTCCAGGCGCAACACCTATGCCCTGGA 1632
256 uAsnPhePhe.....GluAspAlaTyrLeuGlyValLeuProL 269
1633 GAACACTATGATCTCAACAGCGCGCGCAACCCGGCGACCCAGCCG. 1681
269 ysPheSerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMet 285
1682 .....ACGTACCTCGAGGCAACATGCTCATCCAGCGAGAGCGAGCG 1726
286 HisProThrGlyAsnValSerTyr..... 293
1727 CAGCGGTGAACCGGTGGGTATCAACGTCGCGGCGCAGATGGCCACCA 1776
293 ..... 293
1777 CAACAGAGCTCCACTGCCCGCGCGACCGCAGCTACAACCTCCAGG 1826
294 .....GlyGluValPheValLysGlnIleTyrAspAlaIleArg 306
1827 AATCGTGGCGGCGAGCGTGTGGATGGAGAGG.....GACGTGTACCTC 1870
307 GlnGlyProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluTh 323
1871 CAAGGACCATCTCGGCCAAG.....ATCCAGAGAC 1902
323 rGlyGlyPheTyr.....AspH 329
1903 GGGGCGGCACTTTCACCCCTCTCCGCCATGGCGGATTCGGATCAAC 1952
329 isValProProLeuAlaValArg.....ProAsnLeu 341
1953 ACCACCGCCCATGATGCTCATCAAGAACACGCGCTGTGCCCGGAATATC 2002
342 ThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAs 358
2003 ACC..... 2005
358 pArgLeuGlyGlyArgMetProThrTrpValIleSerProTyrSerLysL 375
2006 .....AGCTTCTCGGACGTCGCGGTC 2027
375 ysGlyTyrIleGluGlnTyrGlyThrAspProValThr 387
.....
```

```
2028 GCAGCTTCATCACCAGTACAGCACCGGCGAGGTCAACC 2065
seq_name: /cgn2_2/gcgdata/geneseq/geneseq/NA2000.DAT:Z49212
seq_documentation_block:
ID Z49212 standard; DNA; 2307 BP.
XX
AC Z49212:
XX
DT 15-MAR-2000 (first entry)
XX
DE Adeno associated virus AAV5 capsid protein VP1 DNA.
XX
KW Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;
KW promoter; Rep protein; capsid protein; regulation; transcripion;
KW replication; chromosomal integration; tissue tropism; cellular receptor;
KW gene therapy; neutralising antibody; erythroid progenitor cell;
KW transduction; cancer; genetic disease; VP1; ss.
XX
OS Adeno associated virus type 5.
XX
XX Key Location/Qualifiers
XX CDS 105..2279
XX FT /*tag= a
XX FT /product= "AAV5 capsid protein VP1"
XX
XX PN W09961601-A2.
XX
XX PD 02-DEC-1999.
XX
XX PF 28-MAY-1999; 99WO-US11958.
XX
XX PR 28-MAY-1998; 98US-0087029.
XX
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX PI Chlorini JA, Kotin RM;
XX
XX DR WPI: 2000-062707/05.
XX DR P-FSDB; Y58160.
XX
XX PT Adeno-associated virus 5 based vectors and particles, useful for gene
XX therapy -
XX
XX PS Claim 30; Page 82; 91pp; English.
XX
XX CC This sequence represents DNA encoding the VP1 capsid protein of
XX adeno associated virus type 5 (AAV5). The invention relates
XX to vectors comprising a pair of AAV5 inverted terminal repeats
XX (ITRs) with a promoter between the ITRs. The vector may comprise
XX the viral genome, or subregions thereof, including sequences
XX encoding Rep proteins and capsid proteins, and is encapsidated
XX in an AAV5 particle. The non-structural Rep proteins Rep40
XX (Y58613), Rep52 (Y58168), Rep68 (Y58164) and Rep78 (Y58159) are
XX involved in regulation of replication and transcription, in addition to
XX the production of progeny genomes. Rep68 and Rep78 are also associated
XX with the stable integration of the viral genome into human chromosomes.
XX The three types of capsid protein VP1 (Y58160), VP2 (Y58161) and VP3
XX (Y58162) assemble to form an icosahedral capsid, and differ from each
XX other by the use of alternative splicing and an unusual translation
XX initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2
XX capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are
XX likely to utilise distinct cellular receptors and are serologically
XX distinct. In a gene therapy application, therefore, AAV5 would allow for
XX transduction of a patient who already possess neutralising antibodies
XX either as a result of natural immunological defence or from prior
XX exposure to AAV2 vectors. The vectors may be useful for transducing
XX erythroid progenitor cells or cells lacking heparin sulphate
XX proteoglycans, which is very inefficient with AAV2-based vectors.
XX The vectors may also be useful for transducing cells with a nucleic
XX acid of interest in order to produce cell lines that could be used to
XX screen for agents that interact with the gene product of the nucleic
XX acid of interest. In addition to transduction of other cell types,
```

CC transduction of erythroid cells would be useful or the treatment of  
 CC cancer and genetic diseases which can be corrected by bone marrow  
 CC transplants using matched donors.

XX  
 SQ Sequence 2307 BP; 587 A; 731 C; 603 G; 386 T; 0 other;

# alignment\_scores:

Quality: 107.00 Length: 346  
 Ratio: 0.738 Gaps: 16  
 Percent Similarity: 41.908 Percent Identity: 19.075

# alignment\_block:

US-09-426-072-2\_copy\_20\_464 x 249212 ..

Align seg 1/1 to: 249212 from: 1 to: 2307

```

99 ThrGlyAsnLeuGlu.....Phety 106
|||||.....|
1296 KGGGCAACACTTTGAGTTACTACACTTGAGGAGTGCCCTTCCA 1345
106 rGlyThrTyrProAsnAsnGlyAlaIleAlaSerGlyLysValVala 123
.....|
1346 CTCGAGCTTCGCTCCCGAGTCAAGACCTGTTCAAGCTGGCCCAACCGCTGG 1395
123 laAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeu 139
.....|
1396 TGGACCACTACTTGTACCGCTTCGTGAGCACAAATAAC.....ACT 1436
140 AlaProGluGluAlaThrArgGlnValMetGlyTyrThrGluGluGlu 156
.....|
1437 GCGGAGTCCAGTTCAACAAGAAGCTGGCGGAGATACGCC..... 1478
156 vAlaProThrLeuValAspLeuValAspGluPheThrPheAsnSert 173
.....|
1479 .....AACACCTACAAAACCT 1494
173 rPheSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeu 189
|||||.....|
1495 GGTTC.....CGGGGCCCATGGCGCAACCCAG..... 1523
190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206
|||.....|
1524 ...GGCTGGAACTGGGCTCCGGGTCAACCGC..... 1553
206 rGlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLysGlyValS 223
.....|
1554 .....GCCAGTGTACAGCGCTTCGCCACGACCAATAGG..... 1586
223 erTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAspSerLeu 239
.....|
1587 .....ATGGAGCTCAGGGCGGAGTTACCAGGTGCCCCCGCAGCCGAAC 1631
240 PhePheThrTyrValAsnGlnThrSerArgSerAsnValProValGlu 256
.....|
1632 GGCATGACC.....AACAACTCCAGGCGAGCAACACCTATGCCCTGGA 1675
256 uAsnPhePhe.....GlnAspAlaTyrLeuGlyValLeuProL 269
|||||.....|
1676 GNACACTATGATCTTCAACAGCCAGCGCGGAACCCGGGACACACCC 1724
269 ysPheSerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMet 285
.....|
1725 .....ACGTACCTCGAGGCGCAACATGCTCATCACCAGCGAGCGAGACG 1769
286 HisProThrGlyAsnValSerTyr..... 293
.....|
1770 CAGCGGTGAACCGCTGGCGTACACAGTCGCGGCGCAGATGCCACCAA 1819
293 ..... 293
1820 CAACAGAGGCTCCACACTGCCCCCGGACCGGACGACGTACAACCTCCAGG 1869

```

```

294 .....GlyGluValPheValLysGlnIleTyrAspAlaIleArg 306
1870 AATCTGTGCGCGCAGCGTGTGGATGGAGAG.....GACGTGTACCTC 1913
307 GlnGlyProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluTh 323
|||||.....|
1914 CAAGGACCCATCTGGGCCAAG.....ATCCCGAGAGAC 1945
323 rGlyGlyPheTyr.....AspH 329
|||||.....|
1946 GGGGGCCACTTTCACCCCTCTCCGGCCATGGCGGATTCGACTCAAC 1995
329 isValProProProLeuAlaValArg.....ProAspAsnLeu 341
|| |||||.....|
1996 ACCCAGCGCCCATGATGCTCATCAAGAACACGCTGTGCGCGGAATATC 2045
342 ThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAs 358
|||||.....|
2046 ACC..... 2048
358 pArgLeuGlyGlyArgMetProThrTrpValIleSerProTyrSerLysL 375
.....|
2049 .....AGTTCTCGGACGTGCCCTCA 2070
375 ysGlyTyrIleGluGlnTyrGlyThrAspProValThr 387
.....|
2071 GCAGCTTCATCACCACGACGACGCGGCGGAGGTCAAC 2108

```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT:249209

seq\_documentation\_block:

ID Z49209 standard; DNA; 4652 BP.

XX Z49209;

XX 15-MAR-2000 (first entry)

XX Adeno associated virus AAV5 genome.

XX Adeno associated virus; AAV5; AAV2; inverted terminal repeat; ITR;  
 KW promoter; Rep protein; capsid protein; regulation; transcription;  
 KW replication; chromosomal integration; tissue tropism; cellular receptor;  
 KW gene therapy; neutralising antibody; erythroid progenitor cell;  
 KW transduction; cancer; genetic disease; ss.

XX Adeno associated virus type 5.

XX WO9961601-A2.

XX 02-DEC-1999.

XX 28-MAY-1999; 99WO-US11958.

XX 28-MAY-1998; 98US-0087029.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Chlorini JA, Kotin RM;

XX WPT: 2000-062707/05.

XX P-PSDB; Y58158, Y58159, Y58160, Y58161, Y58162, Y58163, Y58164.

XX Adeno-associated virus 5 based vectors and particles, useful for gene therapy -

XX Claim 10; Fig 4; 91pp; English.

XX This sequence represents the genome of adeno associated virus type  
 CC 5 (AAV5). The invention relates to vectors comprising a pair of AAV5  
 CC inverted terminal repeats (ITRs) with a promoter between the ITRs.  
 CC The vector may comprise the viral genome, or subregions thereof,  
 CC including sequences encoding Rep proteins and capsid proteins, and  
 CC is encapsidated in an AAV5 particle. The non-structural Rep proteins

CC Rep40 (Y58613), Rep52 (Y58168), Rep68 (Y58164) and Rep78 (Y58159) are  
 CC involved in regulation of replication and transcription, in addition to  
 CC the production of progeny genomes. Rep68 and Rep78 are also associated  
 CC with the stable integration of the viral genome into human chromosomes.  
 CC The three types of capsid protein VP1 (Y58160), VP2 (Y58161) and VP3  
 CC (Y58162) assemble to form an icosahedral capsid, and differ from each  
 CC other by the use of alternative splicing and an unusual translation  
 CC initiation codon (in VP2). AAV5 capsid protein is distinct from AAV2  
 CC capsid protein and exhibits different tissue tropism. AAV2 and AAV5 are  
 CC likely to utilise distinct cellular receptors and are serologically  
 CC distinct. In a gene therapy application, therefore, AAV5 would allow for  
 CC transduction of a patient who already possess neutralising antibodies  
 CC either as a result of natural immunological defence or from prior  
 CC exposure to AAV2 vectors. The vectors may be useful for transducing  
 CC erythroid progenitor cells or cells lacking heparin sulphate  
 CC proteoglycans, which is very inefficient with AAV2-based vectors.  
 CC The vectors may also be useful for transducing cells with a nucleic  
 CC acid of interest in order to produce cell lines that could be used to  
 CC screen for agents that interact with the gene product of the nucleic  
 CC acid of interest. In addition to transduction of other cell types,  
 CC transduction of erythroid cells would be useful or the treatment of  
 CC cancer and genetic diseases which can be corrected by bone marrow  
 CC transplants using matched donors.

XX SQ Sequence 4652 BP; 1169 A; 1325 C; 1241 G; 917 T; 0 other;

alignment\_scores:  
 Quality: 107.00 Length: 346  
 Ratio: 0.738 Gaps: 16  
 Percent Similarity: 41.908 Percent Identity: 19.075

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x Z49209 ..

Align seg 1/1 to: Z49209 from: 1 to: 4652

99 ThrGlyAsnLeuGlu.....PheTy 106  
 |||||.....  
 3403 ACGGGACACACTTGTGATTACCTACAACTTTGAGAGGTGCCCTTCCA 3452  
 106 rGlyThrTyrrProAsnGlyAlaIleAlaSerGlyLysValAla 123  
 |||||.....  
 3453 CTCAGCTTCGCTCCAGTCAGAACCTCTTCAAGCTGGCCACCGCTGG 3502  
 123 laAspGlySerGlyPheLeuAsnAlaGlnLeuAsnAspTyrrProLysLeu 139  
 :|||.....  
 3503 TGGACCACTACTTCTACCGCTTCGTGACGACAAATAAC.....ACT 3543  
 140 AlaProGluGluAlaThrArgGlnValMetGlyTyrrThrThrGluGlu 156  
 :|||.....  
 3544 GCGGAGTCCAGTTCAACAAGAACCTGGCGGGAGATACGCC..... 3585  
 156 uValProThrLeuValAspLeuValAspGluPheThrThrPheAsnSerT 173  
 .....  
 3586 .....ACACCTACAAAACCT 3601  
 173 rpPheSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeu 189  
 |||||.....  
 3602 GGTTC.....CCGGGCGCCATGGCGCAACCCAG..... 3630  
 190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206  
 |||||.....  
 3631 ...GGCTGGACCTGGGCTCGGGGTCAACCGC..... 3660  
 206 rGlyIleSerSerIlePheGluAlaAlaAsnGluLysGlyValS 223  
 :|||.....  
 3661 .....GCCAGTTCAGGCGCTTCGCCACGACCAATAGG..... 3693  
 223 exTrpLeuAsnTyrrAspGlyThrAsnGlyGluPheGluProAspSerLeu 239  
 :|||.....  
 3694 ....ATGGAGCTCGAGGCGGAGTTACACAGGTGCGCCCGGACCCGCAAC 3738

240 PhePheThrTyrrValAsnGlnThrSerArgSerAsnValValProValG 256  
 :|||.....  
 3739 GGCATGACC.....AACAACTCCGGGACAGACACCTATGCTCCCTGGA 3782  
 256 uAsnPhePhe.....GlnAspAlaTyrrLeuGlyValLeuProL 269  
 |||||.....  
 3783 GAACACTATGATCTTCAACAGCGACCGCGGAACCGCGGACACCGGCC 3831  
 269 ysPheSerTyrrIleAsnProSerCysGlyThrAsnThrAsnSerMet 285  
 :|||.....  
 3832 .....ACGTACCTCGAGGACACATGCTCATCACCGAGGAGGAGACG 3876  
 286 HisProThrGlyAsnValSerTyrr..... 293  
 :|||.....  
 3877 CAGCGGCTGAACCGCGTGGCTACAACTGCGGCGGAGATGGCCACCAA 3926  
 293 ..... 293  
 3927 CAACCCAGAGCTCCACCACTGCCCGCGGACCGGACGTACACCTCCAGG 3976  
 294 .....GlyGluValPheValLysGlnIleTyrrAspAlaLeuArg 306  
 :|||.....  
 3977 AATCGTCCCGCGGCGAGGTGTGGATGGAGAGG.....GACGTGTACTC 4020  
 307 GlnGlyProGlnTrpAspLysThrLeuLeuPheIleThrTyrrAspGluTh 323  
 |||||.....  
 4021 CAAGGACCCATCTGGGCCAAG.....ATCCACAGAGAC 4052  
 323 rGlyGlyPheTyrr.....AspH 329  
 |||||.....  
 4053 GGGGCGGCACATTTCACCCCTCTCGGCGCATGGCGGATTTCGGACTCAAC 4102  
 329 isValProProLeuAlaValArg.....ProAspAsnLeu 341  
 |||||.....  
 4103 ACCACCGCCGATGATGCTCATCAGACACGCGTGTGCCGGAATATC 4152  
 342 ThrTyrrThrGluThrAlaLysAsnGlyGlnLysTyrrThrLeuHisPheAs 358  
 |||||.....  
 4153 ACC..... 4155  
 358 pArgLeuGlyGlyArgMetProThrTrpValIleSerProTyrrSerLysL 375  
 :|||.....  
 4156 .....AGCTTCTCGGACGTCGCGCTCA 4177  
 375 ysGlyTyrrIleGluGlnTyrrGlyThrAspProValThr 387  
 :|||.....  
 4178 GCAGCTTCATCACCCAGTACAGCCGCGGAGGTGTCACC 4215

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/NA2000.DAT.Z34958

seq\_documentation\_block:

ID Z34958 standard; CDNA; 1718 BP.

XX	AC	Z34958;
XX	DT	28-FEB-2000 (first entry)
XX	DE	Soybean acid triacylglycerol lipase cDNA.
XX	KW	Triacylglycerol lipase; soybean; fatty acid; seed oil;
XX	KW	vegetable oil; transgenic plant; ss.
XX	OS	Glycine max.
XX	XX	
XX	Key	Location/Qualifiers
XX	FT	184..1416
XX	FT	/*tag= a
XX	XX	
XX	PN	W09955883-A2.
XX	XX	
XX	PD	04-NOV-1999.
XX	XX	
XX	PF	29-APR-1999; 99WO-US09280.

30-APR-1998; 98US-0083688.  
(DUPO) DU PONT DE NEMOURS & CO E I.  
Cahoon EB, Cahoon RE, Kinney AJ, Rafalski JA;  
WPI; 2000-062036/05.  
P-PSDB; Y32309.  
Novel plant triacylglycerol lipase polynucleotides used to alter the level of the enzyme in transgenic plants  
Claim 2; Page 48; 65pp; English.  
This is the nucleotide sequence of a contig assembled from the entire cDNA insert in clone ssl.pk0022.al and a portion of the cDNA insert in clone sdp3c.pk004.n3 encoding the entire soybean acid triacylglycerol lipase (TAGL) (see Y32309). The clones were isolated from soybean developing pod and seedling (5-10 day post germination) cDNA libraries. Novel acid and neutral TAGL polypeptides (see Y32301-17) and polynucleotides (see Z34950-66) from corn, Catalpa, rice, soybean and wheat tissues are disclosed. The enzymes may be prepared recombinantly and used to raise antibodies, which are used for detecting the enzymes in situ in cells or in vitro in cell extracts. The polynucleotides may be used to create transgenic plants in which the TAGL levels are present at higher or lower levels than normal, or in cell types or developmental processes where they are not normally found. This would alter the level of triacylglycerol and cholesterol esters found in those cells. Accumulation of fatty acids with unusual structures may be a positive phenotype in plants used for foods. In addition, it may be desirable to eliminate expression of TAGL genes for certain applications. TAGL enzymes may also be useful for the processing of plant seed oils and for the development of novel seed oils. The TAGL enzymes can also be used as targets to facilitate the design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because inhibition of the activity of either of the enzymes could lead to an inhibition of plant growth. The polynucleotides also serve as a source of probes and primers, which are useful for genetic mapping, as markers for traits linked to those genes, and to isolate homologous sequences from other species.

Sequence 1718 BP; 492 A; 353 C; 344 G; 529 T; 0 other;

alignment\_scores:  
Quality: 106.50 Length: 380  
Ratio: 0.609 Gaps: 25  
Percent Similarity: 46.053 Percent Identity: 24.737

alignment\_block:  
US-09-426-072-2\_COPY\_20\_464 x Z34958 ..  
Align seg 1/1 to: Z34958 from: 1 to: 1718

32 ValGluAsnIleValTrp.....Le 38  
|||||.....  
478 GTAGATGATGATGATGCTTAAACCCACCAGCAAGATCTGCCGTT 527

38 uIleLeuGluAsnArgSerPheAspAsnIleLeuGlyGlyValArgArg 55  
|||||.....  
528 GATTTTACGTGATTAATGGATTGACGTGCTGGATTGCAACACAGA.... 573

55 InGlyLeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrLysAsn 71  
|||||.....  
574 .....GGAACACAGATATAGTCGCCGACAC 597

72 AlaSer...AspProSerSerClyLysTyrCysThr..... 82  
|||||.....  
598 ATTCATGTCACCCCTCTAGCCAGCCCTATTGGAAATGGCTCTGGGATGA 647

83 .....GlnAlaLysAspTyrAspSerValPheAsnAspProAspHisServ 98

648 ACTGTGCTCCCTATGATTTCCCGCGGTGTTAAATATGTG...TTACGCC 694  
|||  
98 alThrGlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAsnAsnGly 114  
|||||.....  
695 AAACGGGCGAGAGATCAATTACGTGGCCATTCATTGGGAACCTTGGTA 744  
|||  
115 AlaIleAlaSer.....GlyLysValValAlaAspGlnSerGlyPh 128  
|||||.....  
745 GCTTTGGCATCTCTCGGAGGAATAATTGGTT..... 777  
|||  
128 eLeuAsnAlaGlnLeuAsnAspTyrProLysLeuAlaPro..... 141  
|||  
778 .....ACCCAGCTGAATCAGCAGCCTTCTGTGAGCCCTATAGCCTATT 820  
|||  
142 .....GluGluAlaThrArgGlnValMet 149  
|||  
821 TAGCCACATGATACAGCAGCTGTGTGTGACCCCAAGTCTTTGTT 870  
|||  
150 GlyTyrTyrThrGluGluValProThrLeuValAspGln 166  
|||  
871 GGT.....GAGATCACTACCTCTTCGGCTAGCA...GA 902  
|||  
166 uPheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsp 183  
|||||  
903 ATTT.....AATC 910  
|||  
183 roAsnArgLeuCysAla.....LeuAlaGlyThrAlaAlaGlyHis 196  
|||||  
911 CAAAAGGTTAGCTGTGATGCTTCTTCAGTCTCTCTGCTCACCCCT 960  
|||  
197 GlyLysAsnAspAspPheLeuAsnTyrGlyLysSerLysSerIle 213  
|||||  
961 GGGATAGACTCTATGACTGTGTGACT...GCCTAACTGGTAAATG 1007  
|||  
213 ePheGluAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyT 230  
|||||  
1008 CTGCCTCAATCTCAACTGTTGATCTATTCTTGATGAAT..... 1047  
|||  
230 hrAsnGlyGluPheGluProAspSerLeuPheThrTyrValAsnGln 246  
|||||  
1048 .....GAGCCTCAGTCA..... 1059  
|||  
247 ThrSerArgSerAsnValValProValGluAsnPheGlnAspAlaTy 263  
|||||  
1060 ACATCAACAAGAACATGTTG.....CACTTGCTCAGACTGTTAG 1100  
|||  
263 rLeuGlyValLeuProLysPheSerTyrIleAsnProSerCysGlyT 280  
|||||  
1101 ACTTGGGCGTTGACAAATTCATATATGTGAGACCA..... 1137  
|||  
280 hrAsnThrAsnSerMetHisProThrGlyAsnValSerTyrGlyGluVal 296  
|||  
1138 ..GACTATAACATTATGCAC.....TATGGAGAAATA 1167  
|||  
297 PheValLysGlnIleTyrAspAlaIleArgGlnGlyProGlnTrpAspLy 313  
|||||  
1168 TTT...CCTCCATCTATAAC...CTTCCAACATCCCCAC.....GA 1205  
|||  
313 sThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheTyrAspHis 329  
|||||  
1206 TCTCCCTCTCTTCATTAGCTAT.....GGTGAAGAGATGCAC 1243  
|||  
330 ValProProLeuAlaValArg.....ProAspAsnLe 341  
|||  
1244 TTTGAGATGCTGCTGATGTGAGAAATTTGCTGATAAACTCAAGTTCAT 1293  
|||  
341 uThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheA 358  
|||||  
1294 GATGACAAACAGCGCAGCGTTCAGTTTCATCCAGGAATATGCTCATGCTGA 1343  
|||  
358 sPArgLeuGlyGlyArgMetPro...ThrTrpValIle 369  
|||





```

267 upro.....LysPheSerTyrIleAsnProSerCysC 278
|||||
2175 CCCTGGTGTATCGTATTAGAACTTCAATATATGAC..... 2214
278 ySGLyThrAsnThrAsnSerMetHisProThrGlyAsnValSerTyrGly 294
2214 .....
295 GluValPheValLysGlnIleTyrAspAlaIleArgGlnGlyProGlnTr 311
|||||
2215 .....CAAGTTTAGAT..... 2226
311 pAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheTyrA 328
|||||
2227 .....ATAATAATTGACGACCAATGGTACGATTG 2257
328 sPHis.....ValProProProLeu 334
|||||
2258 TTCATAAATCAGGAGATGTTCCATTATATATAAGATACCAACAGATCT 2307
335 AlavalatgProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyCl 351
|||||
2308 CTATACATGACCGAGGATATCAACTTCTATATATGTTCCGAAACGAAAG 2357
351 nLysTyrThrLeuHis.....PheAspArgLeuGlyGlyArgMetP 365
|||||
2358 AAAACCAATCTAGAGCGTAGAGCGTCGACCGTGTGGTGATCCCATG. 2406
365 rothrTrpValIleSerProTyrSerLysLysGlyTyrIleGluGlnTyr 381
|||||
2407 .....AGGATGGATAGGTAT 2421
382 GlyThr 383
|||||
2422 GSTACC 2427

```

seq\_name: /cgn2\_2/gcdata/geneseq/geneseqn/NA1999.DAT.230163

seq\_documentation\_block:

ID 230163 standard; DNA; 34094 BP.

AC 230163;

26-JAN-2000 (first entry)

Complete nucleotide sequence of the PAV-3 genome.

PAV-3; defective recombinant PAV vector; live recombinant virus;  
 subunit vaccine; nucleic acid immunisation; gene therapy;  
 genetic disease; hemophilia; cystic fibrosis; cancer; viral infection;  
 acquired immune deficiency syndrome; PAV antigen; porcine pathogen; ds.

Porcine adenovirus Type 3.

W09953047-A2.

21-OCT-1999.

15-APR-1999; 99MO-US08220.

15-APR-1998; 98US-0081882.

(UYSA-) UNIV SASRATCHEWAN.

Reddy PS, Tikoo SK, Babiuk LA;

WPI; 1999-620422/53.

New nucleic acids from the genome of porcine adenovirus-3, and derived  
 gene therapy vectors, particularly for immunization

Example 2; Fig 1; 87pp; English.

XX

CC The present sequence represents the complete nucleotide sequence of the  
 CC genome of porcine adenovirus-3 (PAV-3). The specification also describes  
 CC a defective recombinant PAV vector comprising inverted terminal repeats  
 CC (ITR), packaging sequences and at least one heterologous nucleotide  
 CC sequence (II), but lacking E1 functions. The defective vectors replicate  
 CC inefficiently in cells (other than helper cells) so are unlikely to be  
 CC immunogenic. Deletion of the E1 (and optionally other regions) increases  
 CC the size of heterologous insert that can be packaged. The PAV-3  
 CC polynucleotides sequences are used to produce (recombinant or defective)  
 CC vectors that can express heterologous proteins, e.g. for making live,  
 CC recombinant virus or subunit vaccines, for nucleic acid immunisation or  
 CC for gene therapy (e.g. of genetic diseases such as hemophilia or cystic  
 CC fibrosis, cancer, or viral infections, including acquired immune  
 CC deficiency syndrome), also for in vitro expression of recombinant  
 CC antigens (for antibody production), antisense RNA, ribozymes or  
 CC therapeutic proteins. They are also used diagnostically to detect PAV  
 CC antigens and/or nucleic acid. The vectors may be used in human or  
 CC veterinary medicine, but particularly for expressing protective  
 CC determinants of porcine pathogens. Regulatory regions may be used to  
 CC control expression of heterologous genes. Antibodies raised against PAV-3  
 CC polypeptides can also be used for diagnosis (to detect PAV-specific  
 CC antigen).  
 CC  
 SQ Sequence 34094 BP; 6240 A; 11070 C; 10693 G; 6091 T; 0 other;

alignment\_scores:

Quality: 105.00 Length: 330  
 Ratio: 0.719 Gaps: 21  
 Percent Similarity: 44.242 Percent identity: 22.727

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x 230163 ..

Align seq 1/1 to: 230163 from: 1 to: 34094

```

111 ProAsnAsn.....GlyAlaIleAlaSerGlyLyl 120
|||||
19484 CCCACAACTGTCAATATCTAGACCTTAAGGTGAACCTGAGCTGGCAA 19533
120 sVal...ValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspT 136
|||||
19534 AGTTAATACCATCTCTCAAGCAAGTTTGTGGGT..... 19567
136 yrProLysLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrTyr 152
|||||
19568 .....CCTATTGATGAACACCGGGAGACATTAA.....ATT 19600
153 ThrGluGluGluValProThrLeuValAspLeuValAspGluPheThrTh 169
|||||
19601 ACAGAAGAAGAA.....GACGAAGAGAGACCAC 19626
169 rPhe.....AsnSerTrpp 174
|||||
19627 CATCGATCCTTTGTATGAGCCCCAACCCAGCTTGGTCCAGCTCGTGGT 19676
174 heSerCysValProGlyProThrAsn..... 182
19677 CAGACATATACCTTCTGCGACTAGCGGAGCTGGAAGAGTTCTCAACAG 19726
183 .....ProAsnArgLeuCys.....AlaLeuAlaGlyThrAlaAlaGl 195
|||||
19727 ACCACACCGCTCAACCTTGTACGGTTCTTATGCTCTCCGCAAAATAT 19776
195 yHis...GlyLysAsnAspAsp.....PheLeuA 205
|||||
19777 TCACGGTGGCAACCAAGGATGACAGGTTACACCATTTGTACTTTTACAA 19826
205 snTyrGlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLysGly 221
|||||
19827 AC.....AATCCCGCCACCGAAGCGGAGCACTCGAAGAAATGGA 19867
222 Val.....SerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluPr 236

```

```

19868 TAAAGCCAAATGTCCACCTATACCTCAGAGGATGTTGACCTAAAGCACC 19917
19869      :      :      :      :      :      :      :      :      :
236  oAspSerLeuPheThrThrValAsnGlnThr..... 247
19918 AGATACTACTCTGGTCTATGCTGTGAATCAACCCAGGAATTCGCTCAAT 19967
248 .....SerArgSerAsnValValProVal 255
19968 ATGGACTTGGACACACAGCGCGCTCCAAACAGGGCAATATACATCGGC... 20014
256 GluAsnPhePheGlnAspAlaTyrLeuGlyValLeuProLysPheSerTy 272
20015 .....TTCAGGACAACTTTATCGGGCTGTTGTACTACACAGCAA 20055
272 rIleAsnProSerCysGlyThrAsnThrAsnSerMetHisProThrG 289
20056 TGGCAACCAAGGCGATGCTAGCGGTGCTCAGGCTCTCAGCTCAACGCGTGG 20105
289 lyAsnVal.....SerTyrGlyGluValPheVal 298
20106 TCGACCTTGACGACAGGAGATCCGGAACCTAGCTAC...CAGCTCTTCCTC 20152
299 LysGlnIleTyrAspAlaIleArgGlnGlyProGlnTrpAspLysThrIe 315
20153 CATAGCTCTATGACAGGTGCGAGGTACTTTAGCTGTGGACACAGGCCAT 20202
315 uLeuPheIleThrTyrAsp.....GluThrGlyGlyP 326
20203 C.....GATTCCTATGACAGAGATGTCGTGCTGCGGCAACATGGCG 20246
326 heTyrAspHisValProPro.....ProLeuAlaValArgProAsp 339
20247 TGGAGGACGAGATGCCCAACTTTGCTTTCCATCGCGCATCGAGACC 20296
340 AsnLeuThrTyrThrGluThrAlaLys.....AsnGlyGlnLysTyrTh 354
20297 AACAATGACATTTACACAGCTCAAAAGAGTGAGATGGT..... 20335
354 rLeuHisPheAspArgLeuGlyGlyArgMetProThrTrp 367
20336 .....GGCTCAAGAGCCACAACTGG 20356

```

seq\_name: /cgn\_2/gcgdata/geneseq/geneseqn/NA1995.DAT:Q91398

seq\_documentation\_block:

ID Q91398 standard; DNA; 3094 BP.

XX AC Q91398;

XX DT 28-FEB-1996 (first entry)

XX DE Porphyromonas gingivalis protease DNA.

XX KW Protease; periodontal disease; pathogenic microbe; diagnosis; ss.

XX OS Porphyromonas gingivalis.

XX FH Key Location/Qualifiers

FT CDS 61..3036

FT /\*tag= a

XX PN JP07135973-A.

XX PD 30-MAY-1995.

XX PF 15-NOV-1993; 93JP-0307084.

XX PR 15-NOV-1993; 93JP-0307084.

XX PA (SUNR ) SUNTORY LTD.

XX WPI; 1995-227397/30.

DR P-PSDB; R77313.

XX An enzyme originated from a periodontal disease pathogenic microbe -  
PT and an antibody against the enzyme, for the determination of the  
PT progress and activity of the disease

XX PS Claim 1; Pages 5-6; 15pp; Japanese.

XX Q91398 encodes R77313 a P. gingivalis (a periodontal disease  
CC pathogenic microbe) protease. An antibody raised against the  
CC enzyme can be used to diagnose the presence and progress of  
CC a periodontal disease, caused by a pathogenic microbe.

XX Sequence 3094 BP; 893 A; 678 C; 742 G; 781 T; 0 other;

alignment\_scores:  
Quality: 104.50 Length: 521  
Ratio: 0.488 Gaps: 30  
Percent Similarity: 41.075 Percent Identity: 19.962

alignment\_block:  
US-09-426-072-2\_COPY\_20\_464 x Q91398 ..

Align seg 1/1 to: Q91398 from: 1 to: 3094

```

17 TyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValG1 33
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 TAGACGCCGGCGGTTACACACCGGTAGAG.....GAAAAACA 764
33 uAsnIleValTrpLeuIleLeuGluAsnArgSerPheAspAsnIleLeuG 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
765 AAATGTCGTATGATGCTCATCTAGTCCAAAGATGAG.....G 805
50 lyGlyValArgGlnGlyLeuAspAsnProIleAsnAsnGlyProPhe 66
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
806 GAGATATTAAA.....GATTTC 822
67 CysAsnTyrLysAsnAlaSerAspProSerSerGlyLysTyrCysThrG1 83
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
823 GTTGATTGGAAAAACCAACCGCGT.....CTCCGTACCGA 857
83 nAlaLysAspTyrAspSerValPheAsnAspProAspHisSerValThrG 100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
858 GGTGAAA.....GTGGCAGAGATATTGCTTCTCCGTTACAG 895
100 lyAsnAsnLeuGluPheTyr.....GlyThrTyrThrProAsnAsnGly 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
896 CTAATGCTATTCGCGAGTTTCGTTAAGCAAGATACGAGAAAGAGTAAT 945
115 AlaIleAlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAl 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
946 GATTTGACCTATGTTCTTTTGGTGGCGATCAC..... 978
131 aGlnLeuAsnAspTyrPro...LysLeuAlaProGluGluAlaThrArg 147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
979 .....AAAGATATTCCTGCCAAATTAATCTCCGGGATCAATCCGACC 1021
147 lnValMetGly.....TyrTyrThrGlu..... 154
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1022 AGGTATATGGACAAATAGTAGTAATGACCACTACACGAAGTCTTCATC 1071
155 .....GluGluValProThrLeuValAspLe 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1072 GGTGTTTCTCATGTGAGACAAAGAGGATCTGAGACACAATTCATCG 1121
163 uValAspGluPhe.....ThrThrPheAsnSerTrpPheSerC 176
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1122 GACTATTCTACTATGAGCGCAATATAACCGAAGACAAATGCTCGGT. 1170
176 ysValProGlyProThrAsnProAsnArgLeuCys.....AlaLeu 189
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1171 .....CAGGCTCTTTGTATTGCTTCGGCTGAA 1197

```

```

190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206
    |||
1198 GGAGGCCCATCCGCAGACATGGTGAAGTGATATCCAGCATGAGAATGT 1247
    |||
206 f..... 206
1248 AATGCCCAATCTGTTACCCAGTATGGCTATFACCAAGATTATCAAAATGTT 1297
207 .....GlyLeuSerLysSerIlePheGluAlaAlaAsnGluLys 220
    |||
1298 ATGATCCGGAGTAACCTCTAAACATTATGATGCTTTCAC...GGA 1344
221 GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAs 237
    |||
1345 GGAATCTCGTTGGTCAACTATACGGGCGCAGGTAGGGAACACAGCTTGGG 1394
237 pSerLeuPhePhe.....ThrTyrValAsnGlnThrSerArgSerAsnV 252
    |||
1395 TAGCTCTCAGTTCGGCACCACCTCATGTGAGCAGCTTACCAACAGCAACC 1444
252 aIValProValGluAsnPhePheGlnAspAla..... 262
    |||
1445 AGCTACCG.....TTTATTTTCGACGTAGCTTGTGTGAATGGCGAT 1485
263 TyrLeuGlyValLeuProLysPhe..... 270
    |||
1486 TTCTTATTCAGCATGCTTGTCTGCAGACGCCCTCATCGTGCACAAAA 1535
271 .....SertyrIleAsnP 275
    |||
1536 AGATGTTAAGCCGACAGTACTGTGTCTATCATAGCGTCTACGATCAACC 1585
275 roSer..... 277
    |||
1586 AGCTCTGGGCTTCTCTATGCGCGGCGAGTACGATGAACGAATTCGT 1635
278 CysGlyThrAsnThrAsnSerMetHisProThr...GlyAsnValSerty 293
    |||
1636 TCGCAAAACACCCGAAACCAACATCAAGCGTACTTTCGGTGGTGTCCACAT 1685
293 rGlyGluValPheVal.....LysGlnIleTyrAspAlaIleArg 307
    |||
1686 GAACGGTATGTTGCTATGTTGGTGAAGATATAAAGAGTGGTGAGAAGA 1735
307 lngIleProGlnTrp.....AspLysThrLeuLeuPheIleThr 319
    |||
1736 TGCTCGACATGGAGCTGTGTTTCGGCGACCCCTCGCTGCTGCTGTACA 1785
320 TyrAspGluThrGlyGlyPheThrAspHisValProProProLeuAlaVa 336
    |||
1786 CTT.....GTCCCGACCAAAATGCAGGT 1808
336 l...ArgProAspAsnLeuThrTyrThrGluThr..... 346
    |||
1809 TACGGCTCCGGCTCAGATTAATTTGACGGATGCTTCAGTCAACGTATCTT 1858
347 .....AlaLysAsnGlyGlnLysTyr 353
    |||
1859 GCGATTATATGTTGCTATGCTACCATTTACGCCAATGGAAGATGTTTC 1908
354 .....ThrLeuHisPheAspArgLe 360
    |||
1909 GGTCTCGCATGCTCGAAATGGAACAGCTACAAATCAATCTGACAGGTCT 1958
360 uGlyGlyArgMetPro...ThrTrpValIleSerProTyrSerLysLysG 376
    |||
1959 GACAAATGAAGACGCTTACCTTACAGTACTGTTGGTTACACACAAAGAGA 2008
376 lyTyrIleGluGlnTyrGlyThrAspProValThrGlyLysProAlaPro 392
    |||
2009 CGGTATTATAGACCATCAACACTAAT.....GGTGAGCCTAACCC 2049

```

```

393 Tyr.....SerAlaThrSerValLeuLysThrLeuG1 403
    |||
2050 TACCAGCCCGTTTCCAACTTGACAGCTACAAACGCGAGGTACAGAAAGTAAC 2099
    |||
403 yTyrLeuTrpAsp 407
    |||
2100 GCTCAAGTGGGAT 2112

```

seq\_name: /cgn2\_2/cgdata/geneseq/geneseq/NA1995.DAT.083484

seq\_documentation\_block:

ID Q83484 standard; DNA; 3159 BP.

XX Q83484;

XX DT 21-SEP-1995 (first entry)

XX DE Arg-gingipain-1 gene.

XX KW Arg-gingipain-1; gingivalis; periodontal disease; vaccine;

XX KW arginine-specific protease; ds.

XX OS Porphyromonas gingivalis.

XX PH Key

XX FT CDS 949..3159

XX FT /\*tag= a

XX FT mat\_peptide 1630..3159

XX FT /\*tag= b

XX FT /product= Arg-gingipain-1

XX PN W09507286-A.

XX PD 16-MAR-1995.

XX PF 09-SEP-1994; 94WO-US10283.

XX PR 24-JUN-1994; 94US-0265441.

XX PR 10-SEP-1993; 93US-0119361.

XX PR 21-OCT-1993; 93US-0141324.

XX PA (UTGE-) UNIV GEORGIA RES FOUND INC.

XX PI Barr PJ, Pavloff N, Potempa J, Travis J;

XX DR WPI; 1995-123373/16.

XX DR P-PSDB; R70186.

XX PT DNA encoding Arg-gingipain proteins - used to develop prods. for

XX PT detection, treatment and prevention of periodontal disease

XX PS Disclosure; Page 63-66; 89pp; English.

XX CC A low mol.wt. arginine-specific gingipain (AG-1) and high mol.wt.

XX CC AG-2) were isolated from P. gingivalis strains H66 (ATCC

XX CC 33277) and W50 (ATCC 53973). The sequences of the proteins were

XX CC used to design PCR primers and probes to isolate AG DNA. Lambda

XX CC DASH and lambda ZAP libraries were screened with a probe based on

XX CC amino acids 11-22 of the AG protein to obtain DNA encoding AG-1

XX CC (Q83484) and AG-2 (Q83489).

XX SQ Sequence 3159 BP; 933 A; 688 C; 741 G; 797 T; 0 other;

alignment\_scores:

Quality: 104.50 Length: 521

Ratio: 0.488 Gaps: 30

Percent Similarity: 41.075 Percent Identity: 19.962

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x Q83484 ..

Align seg 1/1 to: Q83484 from: 1 to: 3159

263 TyrLeuGlyValLeuProLysPhe..... 270  
: : : : : : : : : : : : : : : : : :  
2374 TTCATTTCAGCATGCTTGCATCGAAGACCCGTGATGGTGACAAAA 2423  
:  
271 .....SerThyrilleAsnP 275  
: : : : : : : : : : : : : : : : : :  
2424 AGATGGTAAGCCGACAGGTACTGTGCTATCATACGGCTCAGCATCAACC 2473  
:  
275 roSer.....Cys 277  
: : : : : : : : : : : : : : : : : :  
2474 AGTCATTGGGCTTCTCTATGCGCGGCGAGATGAGATGAACGAATTTCTG 2523  
:  
278 CysGlyThrAsnSerMetHisProThr...GlyAsnValSerTy 293  
: : : : : : : : : : : : : : : : : :  
2524 TGGAAAAACACCGAACCAATCAAGCGTACTTTCGGTGGTCCACCAT 2573  
:  
293 rgIyGluValPheVal.....LysGlnIleTyAspAlaIleArgG 307  
: : : : : : : : : : : : : : : : : :  
2574 GAACGGTATGTTTGCTATGGTGGAAAGTATAAAAAGGATGGTGAGAAGA 2623  
:  
307 lngIyProGlnTrp.....AspLysThrLeuLeuPheIleThr 319  
: : : : : : : : : : : : : : : : : :  
2624 TGCTCGACACATGSACTGTGTTTTCGGCGACCCCTCGCTCGTTCGTACA 2673  
:  
320 TyrAspGluThrGlyGlyPheTyAspHisValProProLeuAlaVala 336  
: : : : : : : : : : : : : : : : : :  
2674 CTT.....GTCCCGACCAAATGACGGT 2696  
:  
336 L..ArgProAspAsnLeuThrTyThrGluThr..... 346  
: : : : : : : : : : : : : : : : : :  
2697 TAGCGGTCGGCTCAGATTAAATTTGACGGATGCTTCAGTCAACGTATCTT 2746  
:  
347 .....AlaLysAsnGlyGlnLysTy 353  
: : : : : : : : : : : : : : : : : :  
2747 GCGATTATANGTGGCTATGCTACCTTTCAGCCAATGGAAAGATGTC 2796  
:  
354 .....ThrLeuHisPheaspArgLeu 360  
: : : : : : : : : : : : : : : : : :  
2797 GGTTCTCGCAGTTGTCGAAAATGGAACAGCTACAAATCAATCTGACAGGTCT 2846  
:  
360 uGlyGlyArgMetPro...ThrTrpValIleSerProTySerLysLysG 376  
: : : : : : : : : : : : : : : : : :  
2847 GACAAATGAACACAGCTTACCCCTACAGTAGTGGTTACACAAGAAGA 2896  
:  
376 lyTyrlleuGlnTyrglyThrAspProValThrGlyLysProAlaPro 392  
: : : : : : : : : : : : : : : : : :  
2897 CGGTTATTAAAGACCATCAACACTAAT.....GTGAGCCCTAACCCC 2937  
:  
393 Tyr.....SerAlaThrSerValLeuLysThrLeuGl 403  
: : : : : : : : : : : : : : : : : :  
2938 TACACGCCGTTTCCAACTTGACACTCAACCGCAGGCTCAGAAAGTAC 2987  
:  
403 yTyrlLeuTrpasp 407  
: : : : : : : : : : : : : : : : : :  
2988 GCTCAAGTGGGT 3000

17 TyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValGI 33  
 1615 TAGCAGCCGGGGCTTACACACCGGTAGAG.....GAAACA 1652  
 33 uAsnIleValTrpLeuIleLeuGluAsnArgSerPheAspAsnIleLeuG 50  
 1653 AAATGGTCCTATGATCGTCATCGTAGCCAAAAGATTATGAG.....G 1693  
 50 LxGlyValAlaArgGlnGlyLeuAspAsnProIleAsnAsnGlyProphe 66  
 1694 GAGATATTAA.....GATTC 1710  
 67 CysAsnTyrLysAsnAlaSerAspProSerSerGlyLysTyrCysThrGI 83  
 1711 GTTGATTCGGAANAACACCGCGGT.....CTCCGTACCGGA 1745  
 83 nAlaLysAspTyrSerValPheAsnAspProAspHisSerValThrG 100  
 1746 GGTGAAA.....GTGGCAGAAGATTATTCCTCCGTTACAG 1783  
 100 LysAsnAsnLeuGluPheTyr.....GlyTyrThrProAsnAsnGly 114  
 1784 CTAAATCTATTTACGAGTTCGTTAAGCAGAATACGAGAAAGGTAAAT 1833  
 115 AlaIleAlaSerGlyLysValAlaAspGlnSerGlyPheLeuAsnAl 131  
 1834 GATTTGACCTATGTTCTTTGGTGGCGATCAC..... 1866  
 131 aGlnLeuAsnAspTyrPro.....LysLeuAlaProGluAlaThrArg 147  
 1867 .....AAAGATTTCCTGCCAAATATTCTCCGGGATCAATCCGACC 1909  
 147 LnValMetGly.....TyrTyrThrGlu..... 154  
 1910 AGGTATATGGACAAATAGTAGGTAAATCACCACACTACACGAAGTCTCATC 1959  
 155 .....GluGluValProThrLeuValAspLeu 1633  
 1960 GGTGCTTTCTCATGTGAGAGAAAGAGATCTGAAGACACAAATCGATCG 2009  
 163 uValAspGluPhe.....ThrThrPheAsnSerTrpPheSerC 176  
 2010 GACTATTCACTATGACCGCAATATACCCAGGACACAAATGGCTGGT. 2058  
 176 ysValProGlyProThrAsnProAsnArgLeuCys.....AlaLeu 189  
 2059 .....CAGGCTCTTTGTTATGCTTCGCGTGAA 2085  
 190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspAspPheLeuAsnTy 206  
 2086 GGAGGCCCATCCGACAGCAATGGTGAAGTGATATCCAGCATGAGAATGT 2135  
 206 r..... 206  
 2136 AATCGCCAAATCGTTACCCAGTATGCGTATACCAAGATATCAATGTT 2185  
 207 .....GlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLys 220  
 2186 ATGATCCGGAGTAACCTCCTAAACACATATTATGATGCTTTTCAAC...GGA 2232  
 221 GlyValSerTrpLeuLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAs 237  
 2233 GGAATCTGGTGGTCACTATACGGGCCCGCGGTACGGAACACAGCTGGGG 2282  
 237 pSerLeuPhePhe.....ThrTyrValAsnGlnThrSerArgSerAsn 252  
 2283 TAGCTCTCACTTCGGCACACACTCATGTGAACGAGCTTACCAACGACACC 2332  
 252 alValProValGluAsnPhePheGlnAspAla..... 262  
 2333 AGCTACCG.....TTTATTTTCGACCTAGCTTTGTTGTAATGGCAT 2373





JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu

Seq primer: TAATACGACTCCTACTATAGG

High quality sequence start: 2

High quality sequence stop: 728.

## FEATURES

source  
1..865  
/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Ea0035b08f"  
/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="Fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/note="vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 242 a 205 c 191 g 224 t 3 others  
ORIGIN

## alignment\_scores:

Quality: 303.00 Length: 290  
Ratio: 1.859 Gaps: 12  
Percent Similarity: 56.207 Percent Identity: 32.414

## alignment\_block:

US-09-426-072-2 x BE053545

Align seg 1/1 to: BE053545 from: 1 to: 865

```

196 ValProGlyProThrAsnProAsnArgLeuCysAlaLeuAlaGlyThrAl 212
   :::::::::: :::::::::::::::::: ::::::::::::::
10  TTGCGCGCGTGCAGCGAGCTTAACAGGCTTACGTACACTACGCGACATC 59
   :::::::::: :::::::::::::::::: ::::::::::::::
212 aAlaGlyHisGlyLysAsnAspAspPheLeuAsnTyrGlyIleSerS 229
   :::::::::: :::::::::::::::::: ::::::::::::::
60 GCATGCTGCCATGAGCAACAACACACAGCTTATCGAAGGATTCCTC 109
   :::::::::: :::::::::::::::::: ::::::::::::::
229 eLysSerIlePheGluAlaAlaAsnGluLysGlyValSerTrpLeuAsn 245
   :::::::::: :::::::::::::::::: ::::::::::::::
110 AAAAAGCTATATTGAATCATTTGGAGAGAAATGGATATAGCTTT..... 153
   :::::::::: :::::::::::::::::: ::::::::::::::
246 TyrAspGlyThrAsnGlyGluPheGluProAspSerLeuPhePheThrTy 262
   :::::::::: :::::::::::::::::: ::::::::::::::
154 .....GGGATTTATTATCAATCTTTTCCATCTACGCTTTTACAGGAA 197
   :::::::::: :::::::::::::::::: ::::::::::::::
262 rValAsnGlnThrSerArgSerAsnValValProValGluAsnPhePheG 279
   :::::::::: :::::::::::::::::: ::::::::::::::
198 GCTTAGG.....CATTGAAATATGTGGACAAATTTCCATC 232
   :::::::::: :::::::::::::::::: ::::::::::::::
279 In.....AspAlaTyrLeuGlyValLeuProLys 288
   :::::::::: :::::::::::::::::: ::::::::::::::
233 AATACGACTCTAAGTTTCAAGCGTCACGTGAAGATCGGAAGCTACCAAT 282
   :::::::::: :::::::::::::::::: ::::::::::::::
289 PheSerTyrIleAsnProSerCysCysGlyThrAsnThr.....As 302
   :::::::::: :::::::::::::::::: ::::::::::::::
283 TATGTGGTGATTTAGGCCAGATATTTTGACATTTTAAACAGCTGTCGAAA 332
   :::::::::: :::::::::::::::::: ::::::::::::::
302 nSerMetHisProThrGlyAsnValSerTyrGlyGluValPheValLysG 319
   :::::::::: :::::::::::::::::: ::::::::::::::
333 CGACGACCACCCCTCCCATGACGTCTCAGAGGCCAGAGCTTGTGAAGG 382
   :::::::::: :::::::::::::::::: ::::::::::::::
319 InIleTyrAspAlaIleArgGlnGlyProGlnTrpAspLysThrLeuLeu 335
   :::::::::: :::::::::::::::::: ::::::::::::::
383 AAATCTATGAAGCGCTCAGATCAAGTCTCAGTGGGAATGAATCTTGTTC 432

```

```

336 PheIleThrTyrAspGluThrGlyGlyPheTyrAspHisValProProPr 352
   :::::::::: :::::::::::::::::: ::::::::::::::
433 CTGGTTCATATATGATGAACATGGTGGTTCTATGACCATGTTCCGACACC 482
   :::::::::: :::::::::::::::::: ::::::::::::::
352 oLeuAlaVal...ArgProAsnAsnLeuThrTyrThrGluThrAlaLysA 368
   :::::::::: :::::::::::::::::: ::::::::::::::
483 AACCGGAGTCCCTAGCCCGCATGATATGTGCGTCTCTGAG..... 522
   :::::::::: :::::::::::::::::: ::::::::::::::
368 snGlyGlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetPro 384
   :::::::::: :::::::::::::::::: ::::::::::::::
523 .....CCTATAACTTCAAGTTTGCCTCTTGGTTGCAGGGTTCCT 564
   :::::::::: :::::::::::::::::: ::::::::::::::
385 ThrTrpValIleSerProTyrSer..... 392
   :::::::::: :::::::::::::::::: ::::::::::::::
565 GCCATTATGGTTTCCCTTGAATTGAGCCTGGACAGTGTGTGCATAGGC 614
   :::::::::: :::::::::::::::::: ::::::::::::::
393 .....LysIleGlyIleGluGlnTyrGlyThrAspProValThrG 407
   :::::::::: :::::::::::::::::: ::::::::::::::
615 CATCAGGCCAGATTCTACATC.....A 637
   :::::::::: :::::::::::::::::: ::::::::::::::
407 lYlAspAlaProTyrSerIleThrSerValLeuLysThrLeuGlyTyr 423
   :::::::::: :::::::::::::::::: ::::::::::::::
638 MAGNTCAGGACATTCCTNCATTCGACGACACTTTAGAG..... 675
   :::::::::: :::::::::::::::::: ::::::::::::::
424 LeuTrpAspIleGluAspPhe...ThrProArgValAlaHisSerProSe 439
   :::::::::: :::::::::::::::::: ::::::::::::::
676 ATTTTCATCTCAAGATTTTCTAACCAAGCGTGATGCTGGCTGGGTC 725
   :::::::::: :::::::::::::::::: ::::::::::::::
439 rPheAspHisLeuIleGly.....ThrThrLeuArgGluAspAlap 453
   :::::::::: :::::::::::::::::: ::::::::::::::
726 CTTTGAT.....ATTGTGGCAATCGAGGACCCCAACAGACTGTC 769
   :::::::::: :::::::::::::::::: ::::::::::::::
453 roileAlaLeuLysThrPro 459
   :::::::::: :::::::::::::::::: ::::::::::::::
770 CCGAAAAACTGGCAAAACCC 789

```

seq\_name: gb\_est48:AW726682

seq\_documentation\_block:

LOCUS AW726682 698 bp mRNA EST 20-APR-2000  
DEFINITION GA\_Ea0022G10 Gossypium arboreum 7-10 dpa fiber library Gossypium  
arboreum cDNA clone GA\_Ea0022G10, mRNA sequence.

ACCESSION

AW726682

VERSION

AW726682.1 GI:7624221

KEYWORDS

EST.

SOURCE

ORGANISM

Gossypium arboreum.

Gossypium arboreum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eutosids II;

Malvales; Malvaceae; Gossypium.

1 (bases 1 to 698)

Leslie, A., Frisch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A.

An integrated analysis of the genetics, development, and evolution

of the cotton fiber

Unpublished (2000)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

High quality sequence stop: 698.

Location/Qualifiers

1..698

/organism="Gossypium arboreum"

/strain="AKA"

/cultivar="8400"

/db\_xref="taxon:29729"

/clone="GA\_Ea0022G10"

/clone\_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue\_type="Fibers isolated from bolls harvested 7-10

dpa"

FEATURES  
source

```

/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      196 a   155 c   156 g   187 t   4 others
ORIGIN

alignment_scores:
  Quality: 286.00      Length: 220
  Ratio: 2.119        Gaps: 9
  Percent Similarity: 61.364      Percent Identity: 34.091

alignment_block:
US-09-426-072-2 x AW726682 ..
Align seg 1/1 to: AW726682 from: 1 to: 698

255 ProAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSerAsnVa 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 CCATCTACGCTTTTACAGGAAGCTTAGG.....CACTT 63

271 lValProValGluAsnPheGln.....AspA 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 GAATATGTGGACAAATTCATCAATACGATCTAAGCTTCAAGCGTCACT 113

281 lAtyrlLeuGlyValLeuProLysPheSerTyrIleAsnProSerCysCys 297
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 GTAAGGATGGGAAGCTACCAAAATATGTGGTATTGAGCCAGATATTTT 163

298 GlyThrAsnThr.....AsnSerMetHisProThrGlyAsnValSe 311
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 GACATTTTAAACAGCTGTGCAACAGCAGCACCCCTTCCCATGACGCTCTC 213

311 rTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgGlnGlyP 328
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 AGAGGCCAGAGAGCTGTGAAGAAATCTATGAAGCGCTCAGATCAAGTC 263

328 roGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyGly 344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 CTGAGTGGAAATGAATCTGTTCTGGTCATATATGATGAACATGGTGGT 313

345 PheTyrAspHisValProProProLeuAlaVal...ArgProAspAsnLe 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 TTCTATGACCATGTTCGACACCAACCGAGTCCCTAGCCCGCATGATAT 363

360 uThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheA 377
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 TGTGCTGCTGAG.....CCTTATACTTCAAGTTTG 395

377 spArgLeuGlyGlyArgMetProThrTrpValIleSerProTyrSerLys 393
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
396 ATGCTCTGGTTGCAGGGTCTCTGCCATATATGTTTCCCTGGATTGAG 445

394 LysGlyTyrIle.....GluGlnTyrGlyThrAspProValThrGlyLy 408
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
446 CTGGAAACAGTGTGCATAGGCCATCAGGCCAGATCCTACA..... 487

408 sProAlaProTyrSerAlaThrSerValLeuLysThrLeuGlyTyrLeuT 425
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
488 ....TCAGATTCGAGCATTCCTCCATTGCGACCAACACTTAAGAAGATT 533

425 rAspIleGluAspPhe...ThrProArgValAlaHisSerProSerPhe 440
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
534 TCATCTCAAGATTTCTAACAAAGCGTGTGCGGTGGTGGTTCCTTT 583

441 AspHisLeuIleGly...ThrThrLeuArgGluAspAlaProIleAlaLe 456
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
584 GATATTGTTGTAATGAAGCAGCCCAAGAACAGACTGTCCAGAAACT 633

456 uLysThrPro 459
|||
634 GGCAGAGCCA 643

seq_name: gb_est70:BE404795

```

```

seq_documentation_block:
LOCUS      BE404795          617 bp      mRNA          EST          21-JUL-2000
DEFINITION WHE1205_E03_105Z5 Wheat etiolated seedling root cDNA library
            Triticum aestivum cDNA clone WHE1205_E03_105, mRNA sequence.
ACCESSION  BE404795
VERSION    BE404795.1
KEYWORDS   EST.
SOURCE     Bread wheat.
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
REFERENCE  1 (bases 1 to 617)
AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
TITLE     The structure and function of the expressed portion of the wheat
            genomes
JOURNAL    Unpublished (2000)
COMMENT   Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105959773
            Fax: 5105959818
            Email: oanderson@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Strategene SK primer.
FEATURES             Location/Qualifiers
     1..617
         /organism="Triticum aestivum"
         /cultivar="Chinese Spring"
         /db_xref="taxon:4565"
         /clone="WHE1205_E03_105"
         /clone_lib="Wheat etiolated seedling root cDNA library"
         /cissue_type="Root"
         /dev_stage="Five day old etiolated seedling"
         /lab_host="E. coli SOLR"
         /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
            Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
            , germinated and grown aseptically in the dark at room
            temperature on filter paper with water, nystatin and
            cefotaxime in covered crystallization dishes. Roots were
            harvested. The tissue, total RNA, and poly(A) RNA were
            prepared, a cDNA library was made, and the cDNA clones
            were in vivo excised to give phagescript phagemids in the
            TJ Close lab (Choi, Close, Fenton) at the University of
            California, Riverside. Plasmid DNA preparations and DNA
            sequencing were performed in the OD Anderson lab (all
            other authors)."
BASE COUNT      141 a   202 c   151 g   123 t
ORIGIN

alignment_scores:
  Quality: 282.50      Length: 216
  Ratio: 2.108        Gaps: 8
  Percent Similarity: 62.037      Percent Identity: 33.796

alignment_block:
US-09-426-072-2 x BE404795 ..
Align seg 1/1 to: BE404795 from: 1 to: 617

221 AspPheLeuAsnTyrGlyIleSerSerLysSerIlePheGluAlaAs 237
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 GACCTCATCAAC...GGCTTCCCGCAGGAAGACCATCTTCGACAGCGCTGA 68

237 nGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheG 254
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 GGAGGAGCGGCTCTCTCTTC.....GGCATCTACTACCAGAAC 106

```



```

254 luProAspSerLeuPhePheThrThrValAsnGlnThrSerArgSerAsn 270
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 TCCGGCCAGGCTCTCTACAGAGGCTCCGAGGCTCAAGCAC..... 150
271 valValProValGluAsnPhePheGln.....As 280
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 .....CTCCTCAAGTTCACAGCAGTACAGCCTCAGTCCAAAGCTCGA 191
280 pAlaTyrLeuGlyValLeuProLysPheSerTyrIleAsnPro..... 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192 CCGCCGAGGCGCAAGCTGCCATTAAGTCTGCTCATCGAGCAGAGTACT 241
295 ..SerCysCysGlyThrAsnThrAsnSerMethHisProThrGlyAsnVal 310
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 TCGATTGCAAGAGTTTCGGGCCAACGACGACCCGCTCGCAGCAGTCT 291
311 SerTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgGlnG 327
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 GCCAGGGCCAGAGTTTCAAGGAGGCTCAGGACACTCGCGGCGAG 341
327 yProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGly 344
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 CCGCAGGTGAAGCAGAGGCTCTCATCATCACTATGATGATGATGGT 391
344 lyPheTyrAspHisValProProProLeuAlaVal...ArgProAspAsn 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 GCTTCTATGACCATGTTCCAGCGCTCTCAAGTGCCTCAGCCTGATGG 441
360 LeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPh 376
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
442 ATCATTT.....GCCCTGACCTTACTACTTCAAGTT 473
376 eAspArgLeuGlyArgMetProThrTrpValIleSerProTyrSerL 393
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
474 TGATCGCTCGGGTGGCGTGCACCGCTTCCTCATCTCGCCCTGGGTG 523
393 ysLysGlyTyrIleGluGlnTyrGlyThrAspProValThrGlyLysPro 409
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 AGAAGGCGACCGTGATGATCAACCAATGGTCT.....GAGGAAC 567
410 AlaProTyrSerAlaThrSerValLeuLysThrLeuGlyTyrLeuTrp 425
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
568 TCACAAATATGAGCATTCATCCATCCCTCGCAGCATTAAGAGCTATTT 615

```

seq\_name: gb\_est47:AW671667

```

seq_documentation_block:
LOCUS      AW671667          504 bp      mRNA      EST      19-JUL-2000
DEFINITION LGI_349_G03.bl_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA
sequence.

```

```

ACCESSION  AW671667
VERSION    AW671667.1  GI:7535561
KEYWORDS   EST
SOURCE     Sorghum.

```

```

ORGANISM   Sorghum bicolor

```

```

REFERENCE  Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Sorghum.
AUTHORS    1 (bases 1 to 504)
TITLE      Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
JOURNAL    An EST database from Sorghum: light-grown seedlings

```

```

COMMENT    Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 500

```

```

FEATURES             Location/Qualifiers
     source            1..504
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LGI)"
                     /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                     seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                     ; the library was made from poly-A RNA in the cloning
                     vector lambda Zap II. Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT          101 a      175 c      139 g      89 t
ORIGIN

```

## alignment\_scores:

```

Quality: 278.50      Length: 168
Ratio: 2.401         Caps: 7
Percent Similarity: 69.048      Percent Identity: 37.500

```

## alignment\_block:

US-09-426-072-2 x AW671667 ..

Align seg 1/1 to: AW671667 from: 1 to: 504

```

68 LeuGlyGly...ValArgGlnGlyLeuAspAsnProIleAsnAsnG 83
||||| ||| |||:|||||:|||||:|||||:|||||:|||||:
13 CTTGGGGGCTCTCTTACCGCAACGACATCGAGGGTCTGGTCGAGCG... 60
83 yProPheCysAsnTyrLysAsnAlaSerAspProSerSerGlyLysTyrC 100
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 .GAGTTCTTCAACCCCTGTCAACATCAGCGAGCGCTCCCAAGGTCAGATCT 109
100 ysThr.....GlnAlaLysAspTyrAspSerValPheAsnAspProAsp 114
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110 GCGCGCGCTCAGCAGGCAACCAACCTCGACCT.....CAGGATCCCAAC 153
115 HisSerValThrGlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAs 131
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 CACAGCGTCAAGCGGCGCAACCTCGAGCTGTTCCAGCACCTTCCACCCGA 203
131 nAsnGlyAlaIleAlaSer.....GlyLysValValAlaAla 143
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 CGAGGCGCAGCTGCGCGGCAACGACCATGCCCTGGCCGG...TACCAGA 250
143 spGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeuAla 159
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 ACATGGGCGCTTTCGTCGAGGAGCAGCTCGCGTGTGTGTAC...AAGACGACA 297
160 ProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGluGluGluVa 176
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 GACTCGGCCAAGCTGCAGAGGTGATGACTACTATCCGCGCCGAGCACAT 347
176 lProThrLeuValAspLeuValAspGluPheThrThrPheAsnSerTrp 193
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 CCGCGCTTCTCGCGCTGGTGGAACTTTTGTCTTTTGACCACTGGT 397
193 heSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeuAla 209
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
398 TCGGCTCTGCGCGGCGGCGCAACCCGAAACCGTGGTGTCTTACCTGACGTCC 447
210 GlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTyrG 226
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
448 GGCACCAAGTCCCGTCATGGCAGCAACGATAACGCTTTTCAACGTTACGG 497
226 yIle 227
|||||
498 ACTC 501

```

seq\_name: gb\_est51:AW929343

```

seq_documentation_block:
LOCUS      AW929343          596 bp      mRNA      EST      30-MAY-2000

```

```

DEFINITION EST38131 tomato flower buds 8 mm to pre-anthesis, Cornell
University Lycopersicon esculentum cDNA clone cT0C7I18 5', mRNA
sequence.
ACCESSION AW929343
VERSION AW929343.1 GI:8104744
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 596)
AUTHORS van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang
F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning
C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.
Generation of ESTs from tomato flower tissue
Unpublished (1999)
TITLE Contact: David Frisch
JOURNAL Clemson University Genomics Institute
COMMENT Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
     source
     1..596
         /organism="Lycopersicon esculentum"
         /cultivar="TA496"
         /db_xref="taxon:4081"
         /clone="cT0C7I18"
         /clone_lib="tomato flower buds 8 mm to pre-anthesis,
         Cornell University"
         /tissue_type="flower"
         /dev_stage="buds 8mm-to-preanthesis"
         /note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
         XhoI; supplier: Tanksley. Flower buds and flowers were
         taken from greenhouse plants (4-8 wks old, TA496). They
         were immediately frozen in liquid nitrogen and then
         size-separated while remaining frozen."
BASE COUNT 172 a 127 c 129 g 168 t
ORIGIN
alignment_scores:
     Quality: 254.50      Length: 188
     Ratio: 2.194        Gaps: 7
Percent Similarity: 61.702 Percent Identity: 33.511
alignment_block:
US-09-426-072-2 x AW929343 ..
Align seq 1/1 to: AW929343 from: 1 to: 596
281 AlaTyrLeuGlyValLeuProLysPheSerTyr..... 291
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 GCTAACTTGCGAAACTTCGAATTCAGTTCGTGATGACAGAGATATTT 51
292 .....lleAsnProSerCysGlyThrAsnThrAsnSerMeth 305
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 CGATGTTAAATTAACCCAGCA.....AATGATGATC 83
305 isProThrGlyValSerTyrGlyGluValPheValLysGlnIleTyr 321
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ATCCATCAGCATGTAGACACTGGACAAAATTTGTTAAGGAGTTTAT 133
322 AspAlaIleArgGlnGlyProGlnTrpAspLysThrLeuPheIleTh 338
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 GAGACTTGGGGCAAGTCACAGTGGAAAGAAATGCACTCTTGATTAC 183
338 rTyrAspGlnThrGlyGlyPheTyrAspHisValProProLeuAla. 354
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 ATATGACGAACATGGTGGCTCTTTGATCATGATACCACTCTGTTCTG 233

```

```

355 .....ValArgProAspAsnLeuThrTyrThrGluThrAlaLysAsnGly 369
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 GTGTGCTTAATCTCTGATGCATTATC.....GGGCTT 265
370 GlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 386
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 CCACCATATTAATCTCAGATTTCACCGCTGGGGGTTCTGTGCCCACTT 315
386 pValIleSerProTyrSerLysGlyTyrIleGluGlnTyrGlyThrA 403
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 CTTGATTTCACCATGGATTGACAGCCCACTGTGATCCATGAGCCGAAAG 365
403 sProValThrGlyLysProAlaProTyrSerAlaThrSerValLeuLys 419
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 GTCCAACTACGTCT.....TCACAAATTGAGCATTTCTTCCATACCA 409
420 ThrLeuGlyTyrIleTrpAspIleGluAsp.....PheThrProArgVa 434
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 ACTGTGAAAGAGCTTTTAACTGGAATCAAGTTTCTTAACCAAGCGAGA 459
434 lAlaHisSerProSerPheAspHisLeu.....IleGlyThrThrLeuA 449
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
460 TCATGCGGCTGGGACATTTGAGAACTATTCCGCTTTCGCTAAGACTCTC 509
449 rgGluAspAlaPro 453
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
510 GTGATGATTGTCCA 523
seq_name: gb_est78:BF004644
seq_documentation_block:
LOCUS BF004644 514 bp mRNA EST 06-OCT-2000
DEFINITION EST43142 KVI Medicago truncatula cDNA clone pkV1-1856, mRNA
sequence.
ACCESSION BF004644
VERSION BF004644.1 GI:10704919
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eutosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 514)
AUTHORS Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and
Fraser, C.M.
VandenBosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L.,
Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and
Fraser, C.M.
TITLE Esrs from roots of Medicago truncatula 24 hours after inoculation
JOURNAL with Sinorhizobium meliloti
COMMENT Unpublished (1999)
Contact: VandenBosch K
Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel: 409 845 7707
Fax: 409 845 2891
Email: kate@bio.tamu.edu
Texas A&M University name: T268454e TIGR sequence name: MFIBH27TKB
More information is available at: http://chryslie.tamu.edu/medicago
Seq primer: SKmod (CTA gAA CTA gtg.gAT CC).
FEATURES
     source
     1..514
         /organism="Medicago truncatula"
         /cultivar="genotype Al7"
         /db_xref="taxon:3860"
         /clone="pkV1-1856"
         /tissue_type="seedling roots"
         /dev_stage="24 hours post-inoculation with Sinorhizobium
         meliloti"
         /lab_host="E. coli strain XL0LR"
         /note="vector: pBluescript SK -; Site_1: EcoRI; Site_2:
         XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

```

was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 148 a 94 c 114 g 158 t  
ORIGIN

alignment\_scores:  
Quality: 253.50 Length: 141  
Ratio: 2.726 Gaps: 2  
Percent Similarity: 65.957 Percent Identity: 38.298

alignment\_block:  
US-09-426-072-2 x BF004644 ..

Align seg 1/1 to: BF004644 from: 1 to: 514

```

277 PhePheCAsAlaTyLeuGlyValLeuProLysPheSerTyrIleAs 293
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 TTCAAGAAAGATGCCAAGAAATGGGAAGCTTCCACCTTTGACACTGATTGA 111
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 nPro.....SerCysCysGlyThrAsnThrAsnSerMetHisProT 307
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 GCCTAGTACTTTGACTTGGCGGTTTGGCGCAATGATGATCACCCTAT 161
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
307 hrGlyAsnValSerTyrGlyValPheValLysGlnIleTyrAspAla 323
   ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
162 CTCATGATGTGCAATGGCAATGTTGGTTAAGGAGATTATGAGACT 211
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 IleArgGlnGlyProGlnTrpAspLysThrLeuLeuPheIleThrTyrAs 340
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
212 CTAAGACCAAGTCCACACTGGAATGAACTCTTTGGTCATTACATATGA 261
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 pGluThrGlyGlyPheTyrAspHisValProProLeuAlaValArgp 357
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
262 TGACATGTTGATGATTTTGGTATCATGTGAAGACTCCTTTTGGT..... 304
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
357 roAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThr 373
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305 .....AACATCTCTAGTCCAGATGGGACACAGGACCTGCTCTTTTTC 349
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
374 LeuHisPheAspArgLeuGlyGlyArgMetProThrTrpValIleSerPr 390
   ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 TTTAAGTTTGTAGGTATAGGTGTTGGGGTGCCTACTATTATGGTCTCTCC 399
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
390 oTyrSerLysGlyTyrIleGluGlnTyrGlyThrAspProValThrG 407
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 TTGGATCAAGAAGGGATTGGTGAAGAGCCCAAGGACGACGTCGCA 449
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
407 lLysProAlaProTyrSerAla 414
   ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 ATTCGAGTTTGACCACTCTTCA 472
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq\_name: gb\_est23:AI691435

seq\_documentation\_block: 631 bp mRNA EST 02-FEB-2000  
LOCUS AI691435  
DEFINITION 606015H11.x1 606 - Ear tissue cDNA library from Schmidt lab Zea  
mays cDNA, mRNA sequence.

ACCESSION AI691435

VERSION AI691435.1 GI:4966579

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.

REFERENCE 1 (bases 1 to 631)

AUTHORS Walbot,V.

TITLE Maize ESTs

from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

# COMMENT

Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 606015 row: H column: 11.

## FEATURES

source  
Location/Qualifiers  
1..631  
/organism="Zea mays"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/clone\_lib="606 - Ear tissue cDNA library from Schmidt lab"  
/tissue\_type="mixed"  
/dev\_stage="ear length from 0.5 cm - 2.0 cm"  
/lab\_host="XL0LR (Stratagene)"  
/note="Organ: Immature ear; Vector: pBK-CMV; Site\_1: EcoRI  
; Site\_2: XhoI; Mixed ear tissue cDNA library from Schmidt lab"

BASE COUNT 168 a 160 c 153 g 150 t  
ORIGIN

alignment\_scores:  
Quality: 253.50 Length: 168  
Ratio: 2.284 Gaps: 5  
Percent Similarity: 66.071 Percent Identity: 35.714

## alignment\_block:

US-09-426-072-2 x AI691435 ..

Align seg 1/1 to: AI691435 from: 1 to: 631

```

281 AlaTyrLeuGlyValLeuProLysPheSerTyrIleAsnPro..... 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
140 GCCAAATGGGAAGCTGCCAAATATATGTGGTGAACAGAGGTACTT 189
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 .SerCysCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 311
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 TGATTGGAGATGTTCCCGACGACAGATGATCATCCCTCGCATGATGG 239
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 erTyrGluValPheValLysGlnIleTyrAspAlaIleArgGlnGly 327
   ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
240 CAAGGGGACAGAGGTTCTGTCAAAGAGGTGTATCAAAACACTGAGAG 289
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 ProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGly 344
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
290 CCGCAGTGGACAGACGCGCTGTATTCACATATGATGAGCATGGTGG 339
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 yPheTyrAspHisValProProLeu.....AlaValArgProAspA 359
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 ATTCATGACCATGTACCTACGCGGTCTGGAGTGGCCCGACGCTGATG 389
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 snLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeu 375
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
390 GAATAGTC.....GGCCTGATCCATACCTACTACTTCAAG 421
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
376 PheAspArgLeuGlyGlyArgMetProThrTrpValIleSerProTyr 392
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422 TTCGAACGGCTTGGGGTGGTGTTCCTCCACCTTTCTCATCTCACC 471
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 rLysLysGlyTyrIleGluGlnTyrGlyThrAspProValThrGly 409
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
472 AGAAGAGGCACATGTGATCCATGACCAATGCTCCA.....CAGATA 515
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
409 roAlaProTyrSerAlaThrSerValLeuLysThrLeuGlyTyrLeu 425
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
516 CCTCGCAATAGCATTCGTCCATCCCTGCCAACACAGTAACAGACT 565
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
426 AspIleGluAsp.....PheThrProArgValAlaHisSerProSer 440
   ||| ::::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pkVI-18E6"
/clone_lib="KVI"
/tissue_type="Seedling roots"
/dev_stage="24 hours post-inoculation with Sinorhizobium
mellioti"
/lab_host="E. coli strain XL0LR"
/notes="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using GigaPack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0LR cells."
BASE COUNT 156 a 101 c 125 g 170 t
ORIGIN

alignment_scores:
  Quality: 242.50      Length: 170
  Ratio: 2.205        Gaps: 5
  Percent Similarity: 64.706      Percent Identity: 34.706

alignment_block:
US-09-426-072-2 x BF004645 ..
Align seg 1/1 to: BF004645 from: 1 to: 552

277 PhePheGlnAspAlaTyrLeuGlyValLeuProLysPheSerTyrIleAs 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 TTCAAGAAAGATGCAAGAAATGGGAAGCTTCACCTTGACAGTGATTGA 111

293 nPro.Ser.....CysCysGlyThrAsnThrAsnSerMetHisPro 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 GCCTGAGTACCTTGACCTGACGGGTTCCGGCAATGATGATCACCA 161

307 ThrGlyAsnValSerTyrGlyGluValPheValLysGlnIleTyrAspAl 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
162 TCTCATGATGTTGCAATGGCAAAATGTTGGTTAAGGAGATTATGAGAC 211

323 allAtrGlnGlyProGlnTrpAspLysThrLeuLeuPheIleThrTyrA 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
212 TCTAAGACAAGTCCACAGTGGAAATGAACCTTTTGGTCATTACATAG 261

340 spGluThrGlyGlyPheTyrAspHisValProProLeuAlaValArg 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
262 ATGAGCATGGTGGATTTTGGATCATGTGAAGACTCCTTTTGT..... 305

357 ProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrTh 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306 .....ACATCTCTAGTCCAGATGGGAACACAGGACCTGCTCTATT 349

373 rLeuHisPheAspArgLeuGlyGlyArgMetProThrTrpValIleSerP 390
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 CTTTAAAGTTTATAGTTAGGTGTTGGGTGCTACTATTATGCTCTCTC 399

390 roTyrSerLysLysGlyTyrIleGluGlnTyrGlyThrAspProValThr 406
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 CTGGATCAAGAAAGGATGTTGGTAAGAGCCCAAGAGCAGGCTGGC 449

407 GlyLysProAlaProTyrSerAlaThrSerValLeuLysThrLeuGlyTy 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 AAT.....TCGGAGTTTGGACACTCTTCAATTCTGCGCCCAAGAA 493

423 rLeuTrpAspIleGluAsp.....PheThrProArgValAlaHisSerP 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
494 GATGTTCAACCTTTCTCTAATCTTTTGGACTCAGAGATGATGGGCTG 543

438 roSerPhe 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 GGACATTT 551

```

```

seq_name: gb_est78:BF010466
seq_documentation_block:
LOCUS BF010466 392 bp mRNA EST 06-OCT-2000
DEFINITION NXCI_084_C01_F NXCI (Nsf Xylem Compression wood Inclined) Pinus
taeda cDNA clone NXCI_084_C01 5', mRNA sequence.
ACCESSION BF010466
VERSION BF010466.1 GI:10710740
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda
REFERENCE 1 (bases 1 to 392)
AUTHORS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
TITLE Sederoff, R.
JOURNAL Molecular Basis of Wood Formation in the Pine Megagenome
COMMENT Unpublished (2000)
Contact: Johnson, Arthur
North Carolina State University
Tel: 919 515 7800
Fax: 919 515 7801
Email: ajohnson@unity.ncsu.edu
Seq primer: T3.
FEATURES
Location/Qualifiers
1..392
/organism="Pinus taeda"
/strain="Coastal plain loblolly pine from North Carolina"
/db_xref="taxon:3352"
/clone="NXCI_084_C01"
/tissue_type="Xylem"
/dev_stage="Juvenile"
/lab_host="XLI-Blue"
/notes="Vector: Bluescript SK; Site_1: Eco RI; Site_2: XhoI
; The library is from early (spring) wood, taken from
three six-year old trees (three different genotypes), in
the juvenile phase. These trees were induced to form
compression wood by bending to a 45 degree angle and tying
them to the ground. Differentiating xylem was harvested
from the bottoms of the inclined stems, and a mixture of
all three genotypes was used for the library. oligo-dT
primed cDNA was directionally cloned into the EcoRI-XhoI
Bluescript SK vector arms. NOTE: The sequences contain a
'cDNA adapter' between the EcoRI site and the start of the
EST. The adapter sequence is 'AATTCGGCAGG'."
BASE COUNT 104 a 93 c 95 g 100 t
ORIGIN

alignment_scores:
  Quality: 241.00      Length: 148
  Ratio: 2.648        Gaps: 5
  Percent Similarity: 61.486      Percent Identity: 37.162

alignment_block:
US-09-426-072-2 x BF010466 ..
Align seg 1/1 to: BF010466 from: 1 to: 392

270 AsnValValProValGluAsnPhePheGlnAspAlaTyrLeuGlyValle 286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10 AATTATGTGGTGTGTCGACGACGCTTACTATGATCTCAAGCTGGC..... 54

286 uProLysPheSerTyrIleAsnProSerCysGlyThrAsnThrAsnS 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 .....CCTGGA.....AATG 64

303 erMetHisProThrGlyAsnValSerTyrGlyGluValPheValLysGln 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 ATGACCATCGTCCCAAGATGTTGGTGAAGGCGCAGAGTTTGTGAAGAG 114

320 IleTyrAspAlaIleArgGlnGlyProGlnTrpAspLysThrLeuLeuPh 336

```

```

115 GTCATGAGTGTGAGATCGAGTCCCGAGTGAACGAGATGTTGTCAT 164
336 eileThrTyAspGluThrGlyGlyPheTyAspHisValProProL 353
165 TATCATACATGAGTACGACGCGGGTTCACGATCATGTGCCAGCGCTG 214
353 eu.....AlaValArgProAspAsnLeuThrTyThrGluThrAlaLys 367
215 TGAAAATGTGCCCTCCAGATGATATCGTTGGATCAGAG..... 255
368 AsnGlyGlnLysTyThrLeuHisPheAspArgLeuGlyGlyArgMetPr 384
256 .....CCCTATTACTTCAATTCGACAGACTGGCGTCCGTGTACC 296
384 oThrTrpValIleSerProTyThrSerLysGlyTyThrIleGluGlnTyG 401
297 GACTATCATGATTTCCCTTGGATCAACAAGGAACAGT..... 336
401 lYThrAspProValThrGlyLysProAlaProTyThrSerAlaThr 415
337 .....GTACATGGCCCAATGACCATATCTCTCTCC 369

```

seq\_name: gb\_est9:AA556789

```

seq_documentation_block:
LOCUS AA556789 661 bp mRNA 28-AUG-1998
DEFINITION 631 Lobloblly pine NA Pinus taeda cDNA clone 2NA9B, mRNA sequence.
ACCESSION AA556789
VERSION AA556789.1 GI:3365803
KEYWORDS EST.
SOURCE lobloblly pine.
  ORGANISM
    Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
    Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE
  1 (bases 1 to 661)
    Riedl,J., Quinell,M., Shoop,E., Swope,K., St.Cyr,S., Carlis,J.,
    Allona,I., Retzel,E., Campbell,M.M., Sederoff,R. and Whetten,R.W.
    Analysis of xylem formation in pine by cDNA sequencing
    Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9693-9698 (1998)
88356220
Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
,NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: F3.
  Location/Qualifiers
    1..661
      /organism="Pinus taeda"
      /strain="Coastal plain lobloblly pine from North Carolina"
      /db_xref="taxon:3352"
      /clone="2NA9B"
      /clone_lib="Lobloblly pine NA"
      /lib_type="Xylem"
      /lab_host="SOLR"
      /note="Vector: lambda-ZAP; Site.1: EcoRI; Site.2: XhoI;
      The result of subtraction of library N with library C.
      Immature xylem from the side of inclined stems of
      differentiating wood was subtracted with immature xylem
      from underside of inclined stems of differentiating
      compression wood. A mixture of four genotypes were used.
      oligo-dT primed cDNA was directionally cloned into the
      EcoRI-XhoI lambda-ZAP vector arms"
    154 a 163 c 158 g 160 t 26 others
BASE COUNT 154 a 163 c 158 g 160 t 26 others
ORIGIN

```

FEATURES

```

source
1..661
/organism="Pinus taeda"
/strain="Coastal plain lobloblly pine from North Carolina"
/db_xref="taxon:3352"
/clone="2NA9B"
/clone_lib="Lobloblly pine NA"
/lib_type="Xylem"
/lab_host="SOLR"
/note="Vector: lambda-ZAP; Site.1: EcoRI; Site.2: XhoI;
The result of subtraction of library N with library C.
Immature xylem from the side of inclined stems of
differentiating wood was subtracted with immature xylem
from underside of inclined stems of differentiating
compression wood. A mixture of four genotypes were used.
oligo-dT primed cDNA was directionally cloned into the
EcoRI-XhoI lambda-ZAP vector arms"
154 a 163 c 158 g 160 t 26 others
BASE COUNT 154 a 163 c 158 g 160 t 26 others
ORIGIN

```

```

alignment_scores:
Quality: 236.50 Length: 174

```

```

Ratio: 2.112 Gaps: 6
Percent Similarity: 64.368 Percent Identity: 32.759
alignment_block:
US-09-426-072-2 x AA556789 ..
Align seg 1/1 to: AA556789 from: 1 to: 661
284 GlyValLeuProLysPheSerTyIleAsnPro.....SerCysCys 297
1 GGGGGGCTGCCAAATCTCGTGTCTATAGANCCCGCTACTTCCATCTCGC 50
297 sGlyThrAsnThrAsnSerMetHisProThrGlyAsnValSerTyGlyG 314
51 GGATTCNCTGCGNAACGATGACCATCTCTCGCATGATGTTTCCAGCGGGC 100
314 luValPheValLysGlnIleTyThrAspAlaIleArgGlnGlyProGlnTrp 330
101 AGAATCTGGTGAAGGAGGTTTACGAGCGGTGANAGCAANCCCGAGTGG 150
331 AspLysThrLeuLeuPheIleThrTyThrAspGluThrGlyGlyPheTyAs 347
151 AACCAACCCCTGCTCTCATATACATGATGAACACGCGGCTTTCTACGA 200
347 pHisValProProLeu.....AlaValArgProAspAsnLeuThrT 362
201 TCATGTGCCAGCCAGTTCGCATATGATGATGATGATGATGATGATGATG 249
362 yrThrGluThrAlaLysAsnGlyGlnLysTyThrLeuHisPheAspArg 378
250 .....GGATCCCTCCCTCTTTTTCATTTTGACAGA 282
379 LeuGlyGlyArgMetProThrTrpValIleSerProTyThrSerLysLysG 395
283 TTAGGCGTGGCGTTCCTCCACTATAATGATGATGATGATGATGATGATG 332
395 yTyIleGluGlnThrGlyThrAspProValThrGlyLysProAlaProT 412
333 AACNGTGGTGCACCGCCCTACGCGGCCA.....TCGCGCGAGCTCAGANT 376
412 yrSerAlaThrSerValLeuLysThrLeuGlyTyThrLeuTrpAspIleGlu 428
377 TCGAGCATTCATCGATTCCTCAACTATCAAGAAATGTTCAATCTCTCT 426
429 Asp.....PheThrProArgValAlaHisSerProSerPheAspHisLe 443
427 GCCAATTTCTCCTCAGANATGCTGGCTGGAGAACTTTTCGAAAGGCT 476
443 uileGlyThrThrLeuArgGlu 450
477 NGTG.....AGCTTGGGGAT 492
seq_name: gb_est42:AW266012
seq_documentation_block:
LOCUS AW266012 601 bp mRNA 29-DEC-1999
DEFINITION L30-2674T3 Ice plant Lambda Uni-Zap XR expression library, 30 hours
NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-2674
5', mRNA sequence.
ACCESSION AW266012
VERSION AW266012.1 GI:6644699
KEYWORDS EST.
SOURCE Common ice plant.
  ORGANISM
    Mesembryanthemum crystallinum
    Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
    Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;
    Caryophyllales; Alzaceae; Mesembryanthemum.
REFERENCE
  1 (bases 1 to 601)
    Cushman, J.C.
    An expressed sequence tag database for the common ice plant,
    Mesembryanthemum crystallinum
    Unpublished (1997)
    Contact: Cushman JC

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2001, 04:51:53 , Search time 2132.85 Seconds

(without alignments)  
4855.975 Million cell updates/sec

Title: US-09-426-072-1\_COPY\_568\_2045

Perfect score: 1478

Sequence: 1 agccctgtcacgtccgagta.....ctccccacaccttttcggta 1478

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estrol:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
 118: em\_estp17:\*  
 119: em\_estp18:\*  
 120: em\_estp19:\*  
 121: em\_estp20:\*  
 122: em\_estp21:\*  
 123: em\_estp22:\*  
 124: em\_estp23:\*  
 125: em\_estp24:\*  
 126: gb\_est58:\*  
 127: gb\_est59:\*  
 128: gb\_est60:\*  
 129: gb\_est61:\*  
 130: gb\_est62:\*  
 131: gb\_est63:\*  
 132: gb\_est64:\*  
 133: gb\_est65:\*  
 134: gb\_est66:\*  
 135: gb\_est67:\*  
 136: gb\_est68:\*  
 137: gb\_est69:\*  
 138: gb\_est70:\*  
 139: gb\_est71:\*  
 140: gb\_est72:\*  
 141: gb\_est73:\*  
 142: gb\_est74:\*  
 143: gb\_est75:\*  
 144: gb\_est76:\*  
 145: gb\_est77:\*  
 146: gb\_est78:\*  
 147: gb\_est79:\*  
 148: gb\_est80:\*  
 149: gb\_est81:\*  
 150: gb\_est82:\*  
 151: gb\_est83:\*  
 152: gb\_est84:\*  
 153: gb\_est85:\*  
 154: gb\_est86:\*  
 155: gb\_est87:\*  
 156: gb\_est88:\*  
 157: gb\_est89:\*  
 158: gb\_est90:\*  
 159: gb\_est91:\*  
 160: gb\_est92:\*  
 161: gb\_est93:\*  
 162: gb\_est94:\*  
 163: gb\_est95:\*  
 164: gb\_est96:\*  
 165: gb\_est97:\*  
 166: gb\_est98:\*  
 167: gb\_est99:\*  
 168: gb\_est100:\*  
 169: gb\_est101:\*  
 170: gb\_est102:\*  
 171: gb\_est103:\*  
 172: gb\_est104:\*  
 173: gb\_est105:\*  
 174: gb\_est106:\*  
 175: gb\_est107:\*  
 176: gb\_est108:\*  
 177: gb\_est109:\*  
 178: gb\_est110:\*  
 179: gb\_est111:\*  
 180: gb\_est112:\*  
 181: gb\_est113:\*  
 182: gb\_est114:\*  
 183: gb\_est115:\*  
 184: gb\_est116:\*  
 185: gb\_est117:\*  
 186: gb\_est118:\*  
 187: gb\_est119:\*  
 188: gb\_est120:\*  
 189: gb\_est121:\*

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	119.6	8.1	667	92	AW565749	AW565749 LG1_349-G
2	74.4	5.0	466	107	BE415670	BE415670 MML037.H0
3	74.4	5.0	472	110	BE586656	BE586656 WHE0509.A
4	73.6	5.0	617	107	BE404795	BE404795 WHE1205.E
5	70.2	4.7	561	135	BE758821	BE758821 au_3066.A
6	63.6	4.3	480	88	AW266416	AW266416 L30-3180T
7	62.6	4.2	601	88	AW266012	AW266012 L30-2674T
8	62	4.2	800	108	BE430717	BE430717 SUN007.E0
9	61	4.1	698	94	AW726682	AW726682 GA_Ea002
10	61	4.1	865	134	BE053545	BE053545 GA_Ea003
11	59.8	4.0	661	9	AA556789	AA556789 631 L0b10
12	58.6	4.0	392	138	BF010466	BF010466 NXCI.084
13	57.2	3.9	631	23	AI691435	AI691435 605015H11
14	56.8	3.8	490	94	AW756470	AW756470 sl21905.Y
15	53.6	3.6	493	134	BE057023	BE057023 945037H04
16	51.8	3.5	581	24	AT731956	AT731956 BNLGH1115
17	51.6	3.5	380	145	T44962	T44962 8225 Lambda
18	51	3.5	487	27	AI994923	AI994923 701500427
19	50.2	3.4	676	19	AI389106	AI389106 GH20192.5
20	48.8	3.3	514	138	BF004644	BF004644 EST433142
21	48.8	3.3	552	138	BF004645	BF004645 EST433143
22	46	3.1	452	87	AW224301	AW224301 EST301028
23	46	3.1	545	108	BE436321	BE436321 EST407399
24	46	3.1	596	97	AW929343	AW929343 EST338131
25	45.8	3.1	400	139	C26175	C26175 C26175 Rice
26	45.6	3.1	500	174	BE7199	BE7199 CP60015B Cp
27	45.2	3.1	470	29	AU095576	AU095576 AU095576
28	44.2	3.0	461	29	AU032411	AU032411 AU032411
29	44.2	3.0	829	107	BE413313	BE413313 MCG012.F0
30	43.8	3.0	466	134	BE051051	BE051051 za71907.b
31	42.8	2.9	373	24	AT737037	AT737037 sb36h07.Y
32	42.8	2.9	774	134	BE040321	BE040321 OE03H03.O
33	42.4	2.9	742	38	AV674403	AV674403 AV674403
34	42.2	2.9	553	87	AW180278	AW180278 MGA0370f
35	42.2	2.9	574	26	AI861086	AI861086 603011G12
36	42	2.8	445	133	BE012244	BE012244 945015H09
37	41.8	2.8	365	169	AZ242255	AZ242255 RPCI-23-9
38	41.6	2.8	457	109	BE516820	BE516820 WHE620.D0
39	41.6	2.8	537	27	AI948319	AI948319 603042C12
40	41.4	2.8	891	192	CNS0416L	AL269814 Tetraodon
41	40.8	2.8	515	109	BE519721	BE519721 HV.CEB001
42	40.6	2.7	681	191	CNS0250D	AL193990 Tetraodon
43	40.4	2.7	436	29	AU068299	AU068299 AU068299
44	40.4	2.7	829	174	B21527	B21527 T30E3-Sp6 T
45	40.2	2.7	495	109	BE498033	BE498033 WHE0951_C

## ALIGNMENTS

## RESULT 1

AW565749  
 LOCUS LG1\_349-G03.gl\_A002 Light Grown 1 (LG1) Sorghum bicolor CDNA, mRNA  
 DEFINITION  
 ACCSSION AW565749  
 VERSION  
 KEYWORDS EST.  
 SOURCE sorghum.

AW565749 667 bp mRNA EST 19-JUL-2000



```

Db 241 AGTGTCCCAATTTACCTCGTCA---TCGAGCAGCGCTACTTCGATTGCAAGAGTCCCG 297
Qy 988 ccaactccatgcacccaccggttaacgtctctctacggttccttcgctcaagcagatct 1047
Db 298 CCAACGACGACACCCGTCGACGAGTTCGACGAGGACAGAGGTTTGTCAAGAGGTCT 357
Qy 1048 atgatccattccagggccctcagtgagggaagaccctgctctcttattcctacagag 1107
Db 358 ACAGAGACGCTGGCGGCGAGCCGCGAGTGGAAACGAGACGGCTCTCATCATCACCTATGATG 417
Qy 1108 agacgggtggtcttacgaccatgtccctccctctctcgc 1148
Db 418 AGCATGGTGCTCTATGACCATGTCCCGACCGCGTCGTC 458

```

## RESULT 3

```

BE586656 472 bp mRNA EST 17-AUG-2000
LOCUS WHE0509_A09_A17R Secale cereale aluminum-stressed root tip cDNA
DEFINITION library Secale cereale cDNA clone WHE0509_A09_A17, mRNA sequence.
ACCESSION BE586656
VERSION BE586656.1 GI:9839688
KEYWORDS EST.
SOURCE rye.
ORGANISM Secale cereale
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Secale.
REFERENCE 1 (bases 1 to 472)
AUTHORS Anderson,O.D., Butler,E., Chao,S., Gustafson,J.P., Han,P.S., Hsia
C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and
Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes - Aluminum-stressed root tip cDNA library from rye (Secale
cereale)
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: M13 reversed primer.
FEATURES
Location/Qualifiers
source 1..472
/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE0509_A09_A17"
/clone_lib="Secale cereale aluminum-stressed root tip cDNA
library"
/tissue_type="Root tip"
/dev_stage="Seedling"
/lab_host="E. coli DH12s"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Seeds
were germinated aseptically on filter paper and
transferred to a hydroponic growth system in a growth
chamber when the primary root was 1 cm in length. After a
2-day establishment period, seedlings were subjected to a
5 ppm aluminum stress prior to tissue harvest. Plants
were grown in an environmental chamber. The tissue, total
RNA, and poly(A) RNA were prepared, and a cDNA library
was made (Butler and Gustafson) at University of Missouri
, Columbia. Plasmid DNA preparations and DNA sequencing
were performed in the OD Anderson lab (all other authors
)."

```

BASE COUNT 94 a 164 c 120 g 94 t  
ORIGIN

```

Query Match 5.0%; Score 74.4; DB 110; Length 472;
Best Local Similarity 65.9%; Pred. No. 1e-10;
Matches 108; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 986 ccaactccatgcacccaccggttaacgtctctctacggttccttcgctcaagcagat 1045
Db 219 CGCCAACGACGACACCCCTCGCAGCAGCTGCCAGGGGCCAGAGTTTGTCAAGAGGT 278
Qy 1046 ctatgatccattccagggccctcagtgagggaagaccctgctcttattcctacacga 1105
Db 279 CTACGAGACGCTGGCGGCGAGCCGCGAGTGGATGAGACTGCTCTCATCATCACCTATGA 338
Qy 1106 ctagaccggtggtcttacgaccatgtccctccctctcgc 1149
Db 339 TGAGCATGGTGCTCTACGACCATGTCCCGACCGCGTCACCG 382

```

## RESULT 4

```

BE404795 617 bp mRNA EST 21-JUL-2000
LOCUS WHE1205_E03_I05ZS Wheat etiolated seedling root cDNA library
DEFINITION Triticum aestivum cDNA clone WHE1205_E03_I05, mRNA sequence.
ACCESSION BE404795
VERSION BE404795.1 GI:9364263
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
REFERENCE 1 (bases 1 to 617)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
Seaton,C.L. and Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Strategene SK primer.
FEATURES
Location/Qualifiers
source 1..617
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1205_E03_I05"
/clone_lib="Wheat etiolated seedling root cDNA library"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Roots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared. a cDNA library was made, and the cDNA clones
were in vivo excised to give pBluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

```

BASE COUNT 141 a 202 c 151 g 123 t  
ORIGIN

Query Match 5.0%; Score 73.6; DB 107; Length 617;



/tissue\_type="Leaf, 30 h 0.4M NaCl"  
 /dev\_stage="Six week old"  
 /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI"  
 BASE COUNT 116 a 108 c 130 g 126 t  
 ORIGIN

Query Match 4.3%; Score 63.6; DB 88; Length 480;  
 Best Local Similarity 55.5%; Pred. No. 1.5e-07;  
 Matches 151; Conservative 0; Mismatches 109; Indels 12; Gaps 1;  
 QY 999 caacccacggtaacgtctctcactcaggtgaggtcttcgtcaagcagatctatgatgccatt 1058  
 Db 4 CACCCTGCCACGAGTGGCGGAGGCGCGTTCGTTAAGGAGGTTTACGAGACGCTG 63  
 QY 1059 gccagggccctcagtgaggacagaccctctcttcattaccctacacgagacgggtggc 1118  
 Db 64 AGCGCGAGTCCGACGTGGAAGAGATGGCGCTGTGTGATCATCTTATGATGAGCATGGTGG 123  
 QY 1119 ttctacgacattgtccctccctctcgcgtccgcccggagacacctgacctacactgag 1178  
 Db 124 TCTATGATCATGTTCCACCCCTGTTCTGGGTGGCTTAACCGGATGGGATCATTTGGG 183  
 QY 1179 actgcgaagaacggtcagaataacactcttcacttcgacctcgtgggtggcgcgcatccg 1238  
 Db 184 CCTG-----AGCGGTTTACTTTTCGTTTCGATCGGTGGGTGGGTGAGGTGCCA 231  
 QY 1239 acctgggttatctccctctacagtaagaagg 1270  
 Db 232 ACCTGCTTATCTCCCTTGGATTGATAGGG 263

RESULT 7  
 AW266012 601 bp mRNA EST 29-DEC-1999  
 LOCUS L30-2674T3 Ice plant Lambda Uni-Zap XR expression library, 30 hours  
 DEFINITION NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-2674  
 5' mRNA sequence.  
 ACCESSION AW266012  
 VERSION AW266012.1 GI:6644699  
 KEYWORDS EST.  
 SOURCE common ice plant.  
 ORGANISM Mesembryanthemum crystallinum  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;  
 Caryophyllales; Aizoaceae; Mesembryanthemum.  
 REFERENCE 1 (bases 1 to 601)  
 Cushman, J.C.  
 An expressed sequence tag database for the common ice plant,  
 Mesembryanthemum crystallinum  
 Unpublished (1997)  
 JOURNAL Contact: Cushman JC  
 COMMENT Department of Biochemistry and Molecular Biology  
 Oklahoma State University  
 350 Noble Research Center, Stillwater, OK 74078-3035, USA  
 Tel: 405-744-6307  
 Fax: 405-744-7799  
 Email: jcushman@biochem.okstate.edu  
 PCR Primers  
 FORWARD: T7  
 BACKWARD: T3  
 Plate: L30-27 row: G column: 6  
 Seq primer: T3  
 High quality sequence stop: 350.  
 Location/Qualifiers  
 1. 601  
 /organism="Mesembryanthemum crystallinum"  
 /db\_xref="taxon:3544"  
 /clone="L30-2674"  
 /clone\_lib="Ice plant Lambda Uni-Zap XR expression library  
 30 hours NaCl treatment"  
 /tissue\_type="Leaf, 30 h 0.4M NaCl"

/dev\_stage="Six week old"  
 /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI"  
 BASE COUNT 155 a 134 c 156 g 156 t  
 ORIGIN

Query Match 4.2%; Score 62.6; DB 88; Length 601;  
 Best Local Similarity 55.4%; Pred. No. 3.1e-07;  
 Matches 150; Conservative 0; Mismatches 109; Indels 12; Gaps 1;  
 QY 1000 accccacggtaacgtctctcactcaggtgaggtcttcgtcaagcagatctatgatgccatt 1059  
 Db 5 ACCCTGCCACGACGTGGCGGAGGCGCGTTCGTTAAGGAGGTTTACGAGACGCTGA 64  
 QY 1060 gccagggccctcagtgaggacagaccctctcttcattaccctacgacgagacgggtgct 1119  
 Db 65 GCGCGAGTCCGACGTGGAAGAGATGGCGCTGTGTGATCATCTTATGATGAGCATGGTGGT 124  
 QY 1120 tctacgacattgtccctccctctcgcgtccgcccggagacacctgacctacactgaga 1179  
 Db 125 TCTATGATCATGTTCCACCCCTGTTCTGGGTGGCTTAACCGGATGGGATCATTTGGG 184  
 QY 1180 ctgcgaagaacggtcagaataacactcttcacttcgacctcgtgggtggcgcgcatgccga 1239  
 Db 185 CTG-----AGCGGTTTACTTTTCGTTTCGATCGGTGGGTGGGTGAGGTGCCA 232  
 QY 1240 acctgggttatctccctctacagtaagaagg 1270  
 Db 233 CCTTGTCTATCTCCCTTGGATTGATAGGG 263

RESULT 8  
 BE430717 800 bp mRNA EST 24-JUL-2000  
 LOCUS SUN007\_E02F991221 ITEC SUN Wheat cDNA Library Triticum aestivum  
 DEFINITION cDNA clone SUN007\_E02, mRNA sequence.  
 ACCESSION BE430717  
 VERSION BE430717.1 GI:9428672  
 KEYWORDS EST.  
 SOURCE bread wheat.  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Triticum.  
 REFERENCE 1 (bases 1 to 800)  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.  
 AUTHORS Anderson O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier  
 S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,  
 Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,  
 Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogilvie, P.,  
 Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,  
 Sorrells, M., Warburton, M. and Wenzel, G.  
 International Triticeae EST Cooperative (ITEC): Production of  
 International Triticeae EST Cooperative (ITEC): Production of  
 Expressed Sequence Tags for Species of the Triticeae  
 Unpublished (2000)  
 JOURNAL Contact: Shariflou M  
 COMMENT Plant Breeding Institute, University of Sydney  
 107 Cobbitty Rd., Cobbitty NSW 2570 Australia  
 Tel: 61 2 9351 8803  
 Fax: 61 2 9351 8851  
 Email: msharif@pop.usyd.edu.au  
 International Triticeae EST Cooperative (ITEC)  
 http://wheat.pw.usda.gov/genome.  
 Location/Qualifiers  
 1. 800  
 /organism="Triticum aestivum"  
 /db\_xref="taxon:4565"  
 /clone="SUN007\_E02"  
 /clone\_lib="ITEC SUN wheat cDNA Library"  
 BASE COUNT 187 a 225 c 198 g 166 t 24 others  
 ORIGIN

Query Match 4.2%; Score 62; DB 108; Length 800;

```

Best Local Similarity 68.3%; Pred. No. 5.1e-07;
Matches 86; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 1032 ttctgaagcagatctatgatccattccgagggccctcagtgaggaagaccctgcgc 1091
Db 510 TTTTGTCAAGGAGGCTTACGAGACACTGGCGGCGAGCCCGCAGTGGAAAGAGAGCGGCTCTC 569

Qy 1092 ttctattaccacagagaccggtgtctctacgacatgtctccctccctcgcgcgc 1151
Db 570 ATCATCACTATGATGAGCATGGTCTCTATGACCATGTTCCACAGCCTGTCAAGTG 629

Qy 1152 cgcgcgc 1157
Db 630 CQTCAG 635

RESULT 9
AW726682 698 bp mRNA EST 20-APR-2000
LOCUS GA_Ea0022G10 Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION arboreum cDNA clone GA_Ea0022G10, mRNA sequence.
ACCESSION AW726682
VERSION AW726682.1 GI:7624221
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Malvales; Malvaceae; Gossypium.
1 (bases 1 to 698)
Leslie, A., Frisch, D., Yu, Y., Wood, T.C., Wing, R.A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
JOURNAL
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 698.
FEATURES
source
1..698
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0022G10"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 196 a 155 c 156 g 187 t 4 others
ORIGIN

Query Match 4.1%; Score 61; DB 94; Length 698;
Best Local Similarity 60.6%; Pred. No. 9.6e-07;
Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 988 ccaactccatgcacccaccggaacgtctctcagcggtgaggtcttcgtaagcagatct 1047
Db 183 CAACAGCAGCACCCCTTCCCATGACGTCTCAGAGGGCCAGAGGCTTGTGAGGAATCT 242

Qy 1048 atgatgcattcccgagggccctcagtgaggaagaccctgcctcttaccacacgacg 1107
Db 243 ATGAAGCGCTCAGATCAAGTCTCTCAGTGAATGAAATCTTGTCTCGTCATATATGATG 302

Qy 1108 agaccggtggtctctacgacatgtccctccctcgcgcgcgc 1152
Db 303 AACATGGTGGTTCTATGACCATGTTCCGACACCAACCGAGTCC 347

Query Match 4.1%; Score 61; DB 94; Length 698;
Best Local Similarity 60.6%; Pred. No. 9.6e-07;
Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 988 ccaactccatgcacccaccggaacgtctctcagcggtgaggtcttcgtaagcagatct 1047
Db 183 CAACAGCAGCACCCCTTCCCATGACGTCTCAGAGGGCCAGAGGCTTGTGAGGAATCT 242

Qy 1048 atgatgcattcccgagggccctcagtgaggaagaccctgcctcttaccacacgacg 1107
Db 243 ATGAAGCGCTCAGATCAAGTCTCTCAGTGAATGAAATCTTGTCTCGTCATATATGATG 302

Qy 1108 agaccggtggtctctacgacatgtccctccctcgcgcgcgc 1152
Db 303 AACATGGTGGTTCTATGACCATGTTCCGACACCAACCGAGTCC 347

```

```

RESULT 10
BE053545 865 bp mRNA EST 08-JUN-2000
LOCUS GA_Ea0035B08f Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION arboreum cDNA clone GA_Ea0035B08f, mRNA sequence.
ACCESSION BE053545
VERSION BE053545.1 GI:8380601
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Malvales; Malvaceae; Gossypium.
1 (bases 1 to 865)
Leslie, A., Frisch, D., Yu, Y., Wood, T.C., Wing, R.A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
JOURNAL
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 728.
FEATURES
source
1..865
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0035B08f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 242 a 205 c 191 g 224 t 3 others
ORIGIN

Query Match 4.1%; Score 61; DB 134; Length 865;
Best Local Similarity 60.6%; Pred. No. 1e-06;
Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 988 ccaactccatgcacccaccggaacgtctctcagcggtgaggtcttcgtaagcagatct 1047
Db 329 CAACAGCAGCACCCCTTCCCATGACGTCTCAGAGGGCCAGAGGCTTGTGAGGAATCT 388

Qy 1048 atgatgcattcccgagggccctcagtgaggaagaccctgcctcttaccacacgacg 1107
Db 389 ATGAAGCGCTCAGATCAAGTCTCTCAGTGAATGAAATCTTGTCTCGTCATATATGATG 448

Qy 1108 agaccggtggtctctacgacatgtccctccctcgcgcgcgc 1152
Db 449 AACATGGTGGTTCTATGACCATGTTCCGACACCAACCGAGTCC 493

RESULT 11
AA556789 661 bp mRNA EST 28-AUG-1998
LOCUS 631 Lobliolly pine NA Pinus taeda cDNA clone 2NA9B, mRNA sequence.
DEFINITION AA556789
ACCESSION AA556789
VERSION AA556789.1 GI:3365803
KEYWORDS EST.
SOURCE lobliolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

```

```

FEATURES
    source
        Fax: 919 515 7801
        Email: ajohnson@unity.ncsu.edu
        Seq primer: T3.
        Location/Qualifiers
            1..392
                /organism="Pinus taeda"
                /strain="Coastal plain loblolly pine from North Carolina"
                /db_xref="taxon:3152"
                /clone="NXCI_084_C01"
                /clone_lib="NXCI (Nsf Xylem Compression wood Inclined)"
                /issue_type="Xylem"
                /cell_type="Compression"
                /dev_stage="Juvenile"
                /lab_host="XL1-Blue"
                /note="Vector: BlueScript SK; Site_1: Eco RI; Site_2: XhoI
                : The library is from early (spring) wood, taken from

```

The library is from early spring wood, taken from three six-year old trees (three different genotypes), in the juvenile phase. These trees were induced to form compression wood by bending to a 45 degree angle and tying them to the ground. Differentiating xylem was harvested from the bottoms of the inclined stems, and a mixture of all three genotypes was used for the library. **oEoR-dT** primed cDNA was directionally cloned into the **oEoR-xhoI** Bluescript SK vector arms. **NOTE:** The sequences contain a 'cDNA adapter' between the **oEoR** site and the start of the EST. The adapter sequence is '**AATTCGCGACGAG**'.

BASE COUNT	104 a	93 c	95 g	100 t
ORIGIN				
Query Match	4.0%	Score 58.6;	DB 138;	Length 392;
Best Local Similarity	52.8%;	Prod. No. 3.9e-06;		
Matches 158;	Conservative 0;	Mismatches 129;	Indels 12;	Gaps 1;

70	CATCTGCTCCAAAGATGTTGGTGAAGAGCGCAAGATTTGTGAAGAAGAGTGATGAGGGTCTCTG	129
1059	cgcaggccctcagtgggacaagaacctgctcttcattacctacgacgacgacgggtggc	1118
130	AGATCGAGTCCCGAGTGGGAACGAGATGTTGTTCTATTACATACATGATGACACGGGGGT	189
1119	tctcagacatgtcctccctcctctcgcgctcgccgacgaacctgacctaacctgag	1178
190	TTCTACGATCATGCCCACGCCCTGT-----GAAATGTCCCTCTCCAGAT	237
1179	actgcgaagaacggtcagaataacactctctcattctaacctggtgggcgcacatccg	1238
238	GATATCGTTGGATCAGGCCCTTATTCTTCAAATTCGACAGACTGGGGCTCCGTGTACCG	297
1239	acctgggttatctccccttcacagtaagaaggatatactcagcagtcaggaacggatcc	1297

RESULT 13

AI691435	631 bp	mRNA	EST
LOCUS	606015H11.x1	606 - Ear tissue	cdna library from Schmidt lab Zea
DEFINITION	maize cDNA, mRNA sequence.		
ACCESSION	AI691435		
VERSION	AI691435		
KEYWORDS	AI691435.1	GI:4966579	
SOURCE	EST.		
ORGANISM	zea mays.		
	zea mays		
	Eukaryota;		
	Viridiplantae;		
	Embryophyta;		
	Tracheophyta;		
	Spermatophyta;		
	Magnoliophyta;		
	Liliopsida;		
	Poales;		
	Poaceae;		
	Zea.		
REFERENCE	1 (bases 1 to 631)		
AUTHORS	Walbot, V.		
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford		
JOURNAL	University		
COMMENT	Unpublished (1999)		
	Contact: Walbot V		

REFERENCE  
Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae; Zea.  
1 (bases 1 to 631)  
Walbot, V.  
Maize ESTs from various cDNA libraries sequenced at Stanford University  
Unpublished (1999)  
Contact: Walbot V.  
JOURNAL COMMENT





Search completed: February 26, 2001, 04:51:57  
Job time: 19502 sec









TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-324-1677  
 TELEFAX: 650-324-1678  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7334 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-928-361B-1

Query Match 3.1%; Score 45.4; DB 3; Length 7334;  
 Best Local Similarity 45.2%; Pred. No. 0.0048;  
 Matches 212; Conservative 0; Mismatches 251; Indels 6; Gaps 1;

Qy 734 caagaatgacgagcttctgaactatggtatctctagcaagtcctcttcgagggccgc 793  
 Db 2594 CAACAACAAACAAACAACTACTACTACTAGCAGACAAACAAACAAACGACAA 2653

Qy 794 caacgagagggcggtcctggctcaactacgagtcgacacgagaggaattcgacgcga 853  
 Db 2654 CAACAACAACTACTACAACTACTACTACTAGCAGACAAACAACTACTACTACGA 2713

Qy 854 ttctctcttctacctaogtcaacacgagacccctccgggtccacggtggtgcggtgaaaa 913  
 Db 2714 CAACAACAAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2773

Qy 914 cttcttccaaagcctcactcgtggtctcctcctcctcctcctcctcctcctcctcctg 973  
 Db 2774 CAACCACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2833

Qy 974 ctggggcaccac 1033  
 Db 2834 CTACAACAAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2893

Qy 1034 cgtcaagcagatctatgatccattcgcagggccctcagtgaggagacacccctgctctt 1093  
 Db 2894 CTACTACCACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2947

Qy 1094 cattacctacgagagacgggtggttcttagacacatgctcctcctcctcctcctcctcctg 1153  
 Db 2948 CACAACAAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3007

Qy 1154 ccgggacacactctacactgagactgagagagacgggtcagaaatac 1202  
 Db 3008 CAACCACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3056

RESULT 7  
 US-08-578-551-2  
 ; Sequence 2, Application US/08578551  
 ; Patent No. 5854050  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dalboge, Henrik  
 ; APPLICANT: Christgau, Stephan  
 ; APPLICANT: Andersen, Lene N.  
 ; APPLICANT: Kofoed, Lene V.  
 ; APPLICANT: Kauppinen, Sakari M.  
 ; APPLICANT: Nielsen, Jack B.  
 ; APPLICANT: Dambmann, Claus  
 ; TITLE OF INVENTION: An Enzyme with Protease Activity  
 ; NUMBER OF SEQUENCES: 34  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 58540500 No. 5854050disk of No. 5854050th America, Inc.  
 ; STREET: 405 Lexington Avenue, 64th Floor  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: United States of America  
 ; ZIP: 10174-6401  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/578,551  
 FILING DATE: 01-FEB-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: DK 0811/93  
 FILING DATE: 06-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO 95/02044  
 FILING DATE: 19-JAN-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lambiris, Elias J.  
 REGISTRATION NUMBER: 33,728  
 REFERENCE/DOCKET NUMBER: 4006, 204-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-867-0123  
 TELEFAX: 212-878-9655  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Aspergillus aculeatus  
 US-08-578-551-2

Query Match 3.1%; Score 45.2; DB 2; Length 1425;  
 Best Local Similarity 45.3%; Pred. No. 0.0027;  
 Matches 164; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

Qy 686 caaccccaacccgctgtgctctgaggaacccgctggtggcatggcgaagaatgacga 745  
 Db 481 CAGCTGGAGCATCTCTTACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540

Qy 746 tgactctctgaactatggtatctcttagaagtcctctcttcgagggcgcaacgagaaggg 805  
 Db 541 GGTACCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

Qy 806 cgtgctctggctcaactacgagtcgacacacgagagaattcgaaacggatctctctctt 865  
 Db 601 CTCCGAGTTCGTCCAGGACACGACGCGGCTGTGGTCTGGCTTCAGCTCGAT 660

Qy 866 cactctacgtaacacagacctcccggtcccaacggtggtcccggttgaaactcttccaaga 925  
 Db 661 CAACACGGTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720

Qy 926 cgcctacctcgggtgctcctccttaaatctctctacattacacccctcctctgctgcgca 985  
 Db 721 CAGCCCTCTCTTCGCGGCTGACCTGAGTACCAGCGCGCGGCTACCTAGGATTTTCGGG 780

Qy 986 caccactctcatgacacccacacggtaacgctctctacggtgaggtctcttcgtaagagat 1045  
 Db 781 CATCGACTCTCCAAGTACACCGGCTCCCTGACCTAGCCCAACGCTCGACGACGCTCC 840

Qy 1046 ct 1047  
 Db 841 CT 842

RESULT 8  
 US-09-190-982-2  
 ; Sequence 2, Application US/09190982  
 ; Patent No. 5998190  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dalboge, Henrik

APPLICANT: Christgau, Stephan  
APPLICANT: Andersen, Lene N.  
APPLICANT: Kofod, Lene V.  
APPLICANT: Kauppinen, Sakari M.  
APPLICANT: Nielsen, Jack B.  
APPLICANT: Dammann, Claus  
TITLE OF INVENTION: An Enzyme with Protease Activity  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 59981900 No. 5998190disk of No. 5998190th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/190,982  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/578,551  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 95/02044  
FILING DATE: 19-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4006.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1425 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: GNA  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Aspergillus aculeatus  
US-09-190-982-2

Query Match 3.1%; Score 45.2; DB 2; Length 1425;  
Best Local Similarity 45.3%; Pred. No. 0.0027;  
Matches 164; Conservative 0; Mismatches 198; Indels 0; Gaps 0;  
QY 686 caaccocaccgttgctgtctggcaggaaacggctgctgggcatggcaagaatgacga 745  
Db 481 CAGGTGGAGCATCTCTACGGCGACGGCAGCTCCGCCAGCGCGAGCTCTACAGGACAA 540  
QY 746 tgacttccgaactatgtatctctagaagtcctatcttgagccgccaagagaagg 805  
Db 541 GGTACCGGTGCGGACGGTGACGGCGACGACGGCGTGGGGCGCCGCCGATCAG 600  
QY 806 cgtgtcctggctcaactacgatggcaccacgagaaattgaaacggattctcttctt 865  
Db 601 CTCGAGTTCGTCCAGGACACCGACACCGCTGTGTGGGTCTGGCTTCAGCTCGAT 660  
QY 866 caactcgtcaacagaccctccggtccaaagtggtgcccgttgaaactcttccaa 925  
Db 661 CACACGGGTCTCCCCGGGGCCAGACACCTTCTTCGACACCGTCAAGTCCAGCTGGA 720  
QY 926 cgtactacctcgtgtctccctaaattcttcttattaaacccctctcgtcggcacc 985

Db 721 CAGCCCCCTCTTTCGCCCTCGACCTGAAGTACCACGCCGCCGGTACCTACGATTTCGGGTT 780  
QY 986 caccacactcatgcaaccaccggtaacgtctctctacggltgaggtctctcgtcaagcagat 1045  
Db 781 CATGACTCTCTCCAGTACACCGGCTCCTCTGACCTACGCCAACGTGACGACTCCAGGG 840  
QY 1046 ct 1047  
Db 841 CT 842  
RESULT 9  
US-08-700-651-1  
; Sequence 1, Application US/08700651B  
; Patent No. 6015882  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN  
; APPLICANT: LEECH, JAMES  
; APPLICANT: NELSON, RICHARD, C.  
; APPLICANT: GUT, JIRI  
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
; TITLE OF INVENTION: INFECTIONS  
; FILE REFERENCE: 480.19-4(HV)  
; CURRENT APPLICATION NUMBER: US/08/700,651B  
; CURRENT FILING DATE: 1997-08-14  
; EARLIER APPLICATION NUMBER: 08/415,751  
; EARLIER FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 5163  
; TYPE: DNA  
; ORGANISM: Cryptosporidium parvum  
US-08-700-651-1

Query Match 3.0%; Score 44.8; DB 3; Length 5163;  
Best Local Similarity 45.6%; Pred. No. 0.006;  
Matches 157; Conservative 0; Mismatches 187; Indels 0; Gaps 0;  
QY 860 cttcttcaactacgtcaaccagacacctcccggtgccacggtggtgcccttgtaaaactctt 919  
Db 600 ctactactactacgactacta 659  
QY 920 ccaagagcctacctcgtgtctctcctaaattcttcttataacccctcctgctgagg 979  
Db 660 ccactactactacgacaa 719  
QY 980 caccacacacactcctgacccccacggtaacgtctctcgtgaggtggtctcgtcaa 1039  
Db 720 caac 779  
QY 1040 gcgagatctatgatgccattccgagggccctcagtgaggacagacccctgctcttctac 1099  
Db 780 caaccac 839  
QY 1100 ctacgagagacgggtggtcttctacgacacatgtccctccctcctcgtcgcgcgga 1159  
Db 840 ccac 899  
QY 1160 caacctgacctacactgagactgcgaagacgggtcagaataca 1203  
Db 900 caaccac 943  
RESULT 10  
US-08-928-361B-4  
; Sequence 4, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,





```

RESULT 14
US-09-096-942-2
; Sequence 2, Application US/09096942
; Patent No. 6027925
; GENERAL INFORMATION:
; APPLICANT: Pollock, Thomas J
; APPLICANT: Mikolajczak, Marcia
; APPLICANT: Yamazaki, Motohide
; APPLICANT: Thorne, Linda
; APPLICANT: Armentrout, Richard W
; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
; FILE REFERENCE: seq list for appl filed from pro. appl
; CURRENT APPLICATION NUMBER: US/09/096,942
; CURRENT FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/049,428
; EARLIER FILING DATE: 1997-06-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 28804
; TYPE: DNA

```











Matches 157; Conservative 0; Mismatches 187; Indels 0; Gaps 0;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



RESULT	10
Z61243	
ID	Z61243 standard; DNA; 2110 BP.
XX	
AC	Z61243;
XX	
DT	30-MAY-2000 (first entry)
XX	
DE	DNA encoding a phenol oxidising enzyme.
XX	
KW	Phenol oxidizing enzyme; fungus; redox reaction; detergent; bleaching;
KW	fabric; pulp; paper; decolourisation; plant-derived food product;
KW	coloured compound; porphyrin; tannin; polyphenol; carotenoid;
OS	anthocyanin; Maillard reaction product; ss.
XX	
OS	Acremonium murorum.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	135..1943
FT	/tag= a
FT	/product= "phenol oxidising enzyme"
XX	
PN	WO200005349-A1.
PD	
PP	03-FEB-2000.
PP	
XX	13-JUL-1999; 99WO-EP04922.
XX	
PR	21-JUL-1998; 98EP-020245A.
XX	
PA	(UNIL ) UNILEVER NV.
PA	(UNIL ) UNILEVER PLC.
PA	(HIND-) HINDUSTAN LEVFE LTD.

```

XX Convents D, Gouka RJ, Van Der Heiden M, Swarthoff T, Verrips CT;
PI WPI; 2000-195101/17.
DR P-PSDB; Y69204.
DR
XX
XX
PT Phenol-oxidizing enzyme from Acremonium, used in detergent compositions
PT for bleaching stains on fabrics
XX
XX
PS Example 2; Page 37-40; 45pp; English.
CC
CC The present sequence encodes a phenol oxidizing enzyme from the
CC fungus Acremonium murorum. The enzyme has the CBS accession number
CC 157.72. The enzyme catalyses redox reactions and is specific for
CC molecular oxygen as the electron acceptor. The phenol oxidising enzyme
CC is specifically used in detergents for bleaching strains on fabrics,
CC but also for bleaching pulp and paper and for decolourisation of
CC plant-derived food products. The enzyme has a pH optimum in the
CC alkaline to neutral range and can bleach a wide variety of coloured
CC compounds, e.g. porphyrins, tannins, polyphenols, carotenoids,
CC anthocyanins and Walliard reaction products.
XX
XX Sequence 2110 BP; 400 A; 732 C; 556 G; 422 T; 0 other;
SQ
Query Match 3.0%; Score 44.2; DB 21; Length 2110;
Best Local Similarity 46.6%; Pred. No. 0.0094;
Matches 142; Conservative 0; Mismatches 163; Indels 0; Gaps 0
OY 780 attctgagcgcgaacgaaggcgtgtctctggtctcaactacgtgcgcacacgga 839
Db 1182 accatcagctccgaacgagccgcggtctgcgcgcctcgatcgaggccaactacgac 1241
OY 840 gaattcgaaacggattctcttcttcaacttcagctcaacagacacctccgcgcacgtg 899
Db 1242 gacacgcgaaggctatgcgtctcaagtcgcgcgagggccctctctctgcgcgcaccc 1301
OY 900 gtgcgcggttgaaactttctccaagacgcctaacctcgggtgctctccctaaattctcttac 959

```











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2001, 00:55:28 ; Search time 4495.24 Seconds  
(without alignments)  
1682.677 Million cell updates/sec

Title: US-09-426-072-1-copy\_568\_2045

Perfect score: 1478

Sequence: 1 agccctgcacgtccgagta.....ctccccaatacttttcggta 1478

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_bal.\*  
2: gb\_ba2.\*  
3: gb\_ov.\*  
4: gb\_ov.\*  
5: gb\_ph.\*  
6: gb\_pl1.\*  
7: gb\_pl2.\*  
8: gb\_pl1.\*  
9: gb\_pr2.\*  
10: gb\_pr3.\*  
11: gb\_ro.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: em\_fun.\*  
15: em\_hum1.\*  
16: em\_hum2.\*  
17: em\_in.\*  
18: em\_om.\*  
19: em\_or.\*  
20: em\_ov.\*  
21: em\_pat.\*  
22: em\_ph.\*  
23: em\_pl.\*  
24: em\_ro.\*  
25: em\_sts.\*  
26: em\_sy.\*  
27: em\_un.\*  
28: em\_vl.\*  
29: gb\_ba3.\*  
30: gb\_in1.\*  
31: gb\_in2.\*  
32: gb\_in3.\*  
33: gb\_pl3.\*  
34: gb\_pr4.\*  
35: gb\_ba1.\*  
36: em\_ba2.\*  
37: em\_htg1.\*  
38: em\_htg2.\*  
39: em\_htg3.\*  
40: em\_htg4.\*  
41: em\_htg5.\*  
42: em\_htg6.\*  
43: em\_htg7.\*

44: em\_htg8.\*  
45: em\_htg9.\*  
46: em\_htg10.\*  
47: em\_hum3.\*  
48: em\_hum4.\*  
49: em\_hum5.\*  
50: em\_hum6.\*  
51: gb\_pr5.\*  
52: gb\_pr6.\*  
53: gb\_pr7.\*  
54: gb\_htg1.\*  
55: gb\_htg2.\*  
56: gb\_htg3.\*  
57: gb\_htg4.\*  
58: gb\_htg5.\*  
59: gb\_htg6.\*  
60: gb\_htg7.\*  
61: gb\_htg8.\*  
62: gb\_htg9.\*  
63: gb\_htg10.\*  
64: gb\_htg11.\*  
65: gb\_htg12.\*  
66: gb\_htg13.\*  
67: gb\_htg14.\*  
68: gb\_htg15.\*  
69: gb\_htg16.\*  
70: gb\_htg17.\*  
71: gb\_htg18.\*  
72: gb\_htg19.\*  
73: gb\_htg20.\*  
74: gb\_htg21.\*  
75: gb\_htg22.\*  
76: gb\_htg23.\*  
77: gb\_sts1.\*  
78: gb\_sts2.\*  
79: gb\_vil.\*  
80: gb\_vil.\*  
81: gb\_pat1.\*  
82: gb\_pat2.\*  
83: em\_htg0.\*  
84: gb\_htg24.\*  
85: gb\_pr8.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	64.8	4.4	82289	7	ATAC009895	AC009895 Arabidops	
	2	64.8	4.4	89934	7	ATAC009327	AC009327 Arabidops	
C	3	55	3.7	96679	7	ATT8E19	AL133315 Arabidops	
	4	52.6	3.6	7218	81	I66494	I66494 Sequence 14	
C	5	52.4	3.5	28291	31	CEF56H9	274473 Caenorhabdi	
C	6	52	3.5	30960	31	CELY46C8AL	AC024785 Caenorhab	
	7	50.6	3.4	2262	1	AF107252	AF107252 Burkholde	
C	8	50.2	3.4	2426	60	AC020448	AC020448 Drosophill	
C	9	50.2	3.4	260910	30	AE003831	AE003831 Drosophill	
	10	49.4	3.3	13669	80	TTV1TP	X14855 Thermoprote	
	11	49.2	3.3	104533	7	F10K1	AC067971 Sequence	
	12	47.8	3.2	4034	80	TTV1VP	X14717 Thermoprote	
C	13	46	3.1	749	80	NPORPT2	D13755 Multicapsid	
	14	46	3.1	30941	29	SC4C6	AL079355 Streptomy	
C	15	46	3.1	131993	80	OPU75930	U75930 Orgyia pseu	
C	16	45.8	3.1	5500	7	ATGLYRP	X58338 A.thaliana	
C	17	45.8	3.1	92620	6	A8026636	A8026636 Arabidops	
C	18	45.6	3.1	144241	60	AC018929	AC018929 Oryza sat	
C	19	45.6	3.1	147371	65	AC027037	AC027037 Oryza sat	
	20	45.4	3.1	5499	31	AF068065	AF068065 Cryptospor	
21	45.4	3.1	108409	79	AF083424	AF083424 Aeteline h		

```

22 45.2 3.1 1425 81 AR068771 Sequence
23 45.2 3.1 1425 81 AR092645 Sequence
24 44.8 3.0 2626 2 PSEPLC
25 44.8 3.0 11622 1 AE004519 Pseudomon
26 44.8 3.0 35284 29 SCE19A
27 44.6 3.0 2783 7 CCLPGLN1
28 44.6 3.0 9958 33 SCYJRI151C
29 44.4 3.0 41906 29 SC5C7
30 44.2 3.0 2126 7 AM0271104
31 44 3.0 750 32 SGSCRHPH
32 44 3.0 1419 33 FSU13050
33 44 3.0 11970 1 AE001978
34 44 3.0 288440 2 CNSPAX04
35 43.8 3.0 2603 7 COGPECA3
36 43.8 3.0 64998 58 AC014035
37 43.8 3.0 269518 30 AE003751
38 43.6 2.9 35284 29 SCE19A
39 43.4 2.9 662 2 MSGMTB40
40 43.4 2.9 3013 2 MSGMTB40A
41 43.4 2.9 4094 2 MTU49511
42 43.4 2.9 31225 2 MTCY98
43 43.4 2.9 40892 81 A59672
44 43.2 2.9 37521 29 SC1A8A
45 43.2 2.9 39446 29 SCE87

```

## ALIGNMENTS

```

RESULT 1
ATAC009895
LOCUS ATAC009895 82289 bp DNA PLN 09-OCT-1999
DEFINITION Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence,
complete sequence.
ACCESSION AC009895.2 GI:6017096
VERSION AC009895
KEYWORDS HTG.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1
AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.,
Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E.,
Bowman, C.B., White, O., Nierman, W.C. and Fraser, C.M.
Mangliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1. (bases 1 to 82289)
Unpublished Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence
2. (bases 1 to 82289)
Lin, X. and Kaul, S.
Direct Submission
Submitted (04-SEP-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org
3. (bases 1 to 82289)
Lin, X.
Direct Submission
Submitted (09-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
e-mail: xlin@tigr.org
On Oct 9, 1999 this sequence version replaced gi:5822970.
Address all correspondence to:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlin@tigr.org
BAC clone T21P5 is from Arabidopsis chromosome III and is near the
molecular marker ml172
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,

```

http://genome.stanford.edu/~chris/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (http://www.tigr.org/tdb/at/at.html). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by RepeatMasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

## FEATURES

```

source
1..82289
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosomes="III"
/map="ml172"
/clone="T21P5"

misc_feature
1..24217
/notes="overlap with BAC clone T12J13
(AC009327:65718..89934)."
```

mRNA

```

complement(join(<8..215,451..587,675..857,959..1050,
1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
2979..3100))
/gene="T21P5.1"
complement(<8..>3100)
/gene="T21P5.1"
/notes="similar to hypothetical protein GB:CA838918
[Arabidopsis thaliana]"
complement(join(8..215,451..587,675..857,959..1050,
1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
2979..3100))
/gene="T21P5.1"
/notes="unknown protein"
/codon_start=1
/protein_id="AAF01600.1"
/db_xref="GI:6017117"
/translation="MSQOLGNNIPLSEVYSLVANKADKESKIRDLPFERSRVEN
YFVKYKVTOLMKFQGNRQKLVKAGLKWKEIGETASIAQIYGHMTWSDGVLK
ESYVYKLTREYKDCGLFQDLINANKQLFLARLWCLVIGRRVHOLVDQFKR
LDECKRTFQIDFKWKVVAQEIYRFLKSDTAPMIRPLRSLVLDNLDAGTPRAS
KSLRUTDALSSYYCNEVYSBELTDSFRMLQCLEWEPGSLYQSGAKGNAPGVG
ARINSQMDPTLPNPKAVLYRPSITFLAVLATICEELPSHGILLLYLSAGKIG
QISSPLSARSATSVENILRDFESHTIKOETPSLQITPSSSVIYPSDLVPFTRKP
LFITIDSSSTVFNKICGAKEGPAALLSPHTPLISADFSRQSGSLFTFLSP
VOAFCLLSIASDMETDIFTKAEKLLSSSNWASTLATSDTLHPVWSQLKDPFLR
RLLRIFECRAVLALYTPVNNKQNPCEPCPSLPLTAPVQSAVFQMANVFGAT
SKFTIPQDITMLESF"
```

mRNA

```

join(<3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)
/gene="T21P5.2"
<3627..>6144
/gene="T21P5.2"
/notes="predicted by genscan"
```

gene

```

join(3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)
/gene="T21P5.2"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF01580.1"
/db_xref="GI:6017057"
/translation="MDDNSESFKRHEIEKDTIASKLEDNTNKLTDPEEMALYAK
VRSQEEIHSIQERIAACLMQDLNRYGKERRKCADLRVAIDKESVTSALNEL
ARRKGLDEENLKLAHDLKIDEDERYIFTMTSLGLLAEGYMPRVANATATSSGIKHLH

```

CDS

```

join(3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)
/gene="T21P5.2"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF01580.1"
/db_xref="GI:6017057"
/translation="MDDNSESFKRHEIEKDTIASKLEDNTNKLTDPEEMALYAK
VRSQEEIHSIQERIAACLMQDLNRYGKERRKCADLRVAIDKESVTSALNEL
ARRKGLDEENLKLAHDLKIDEDERYIFTMTSLGLLAEGYMPRVANATATSSGIKHLH

```



DOLQWTKACNDRIRLSSIVENQPTDFISKDNHDPNRNKTQASYGSTDRGNDYOTN  
 EQLPPNENYTRNPHNIMQDTESLRNFNOIGGSGOIFPQPKRENFYPLSSVAGKE  
 MTQREKARNKSMFDAYNGNEEFASHVYEGPGIDGQIIIGDAIPGKVLGCGFPVR  
 GTCLCMQWVRHLEDCTROYTEGATHPEYIVTADDDVKLIAVEICIPMDQDQGVKYRD  
 FSGLYSNESVSKVDLLIMOGELVRLFANDQNKIRCGNVFNGT"  
 6225..6289  
 /note="exon predicted by xgrail, quality marginal"  
 6384..6503  
 /note="exon predicted by xgrail, quality excellent"  
 6794..6839  
 /note="exon predicted by xgrail, quality excellent"  
 6928..7033  
 /note="exon predicted by xgrail, quality marginal"  
 7293..7411  
 /rpt\_family="(TAAAA)n"  
 complement(7488..7547)  
 /note="exon predicted by xgrail, quality  
 excellent\_shadowexon"  
 complement(7654..7784)  
 /rpt\_family="(TAA)n"  
 complement(7819..7864)  
 /note="exon predicted by xgrail, quality good\_shadowexon"  
 8054..8113  
 /note="exon predicted by xgrail, quality marginal"  
 8263..8380  
 /rpt\_family="(TAAAA)n"  
 complement(8592..8627)  
 /note="exon predicted by xgrail, quality  
 marginal\_shadowexon"  
 <8652..>9527  
 /gene="T21P5.3"  
 <8654..>9527  
 /gene="T21P5.3"  
 /note="contains zinc finger motif, C3HC4 type (RING  
 finger)"  
 8652..9527  
 /gene="T21P5.3"  
 /note="unknown protein"  
 /codon\_start=1  
 /protein\_id="AA01602.1"  
 /db\_xref="GI:6017119"  
 /translation="MGSTGNPNWGTYYVDRDCSQVCSVPCOWCVVFPPPPSFY  
 LODEDSSDFSPLLIALIGILASAFILVSYTLISKCHRRNNSSTSAAINRI  
 SDDYWGQTNNNNGATNPQITGGGGDLSEIKSYIKYKMDGEVESSDCS  
 VCLSEFQENSLRLPKCNHAFVPCIDITLWLSHSCPLCFRAFIYTSAAVEIDLNQ  
 QIVTENNISTGDSVVVNLNLENSRSRNETNEGSTPTXTGDAFKRRRAEIGVVK  
 LRRCGIDRRYTKGD"  
 complement(10838..10896)  
 /note="exon predicted by xgrail, quality  
 marginal\_shadowexon"  
 11174..11254  
 /note="exon predicted by xgrail, quality good"  
 11690..11700  
 /note="exon predicted by xgrail, quality marginal"  
 join(<11814..12256,12381..12578,12676..13085,13162..13332,  
 13418..>13761)  
 /gene="T21P5.4"  
 <11814..>13761  
 /gene="T21P5.4"  
 /note="similar to unknown protein GB:AAC32238 [Arabidopsis  
 thaliana]"  
 join(11814..12256,12381..12578,12676..13085,13162..13332,  
 13418..13761)  
 /gene="T21P5.4"  
 /note="unknown protein"  
 /codon\_start=1  
 /protein\_id="AA01581.1"  
 /db\_xref="GI:6017098"  
 /translation="MAETKGSSEYPIKTIIVLVQENRSFDHTLWFKELNREIDGMV  
 KSKQKPNFGSSDLNHNVYVQSQYVDNPNFHSIRDIYEQVFGPMDSPHPNPG  
 PATMKSFAQNAERKMGKSSAVMNGFKPDALPVYKELVNFACIDRWFASVPGATQPN  
 RLFTSHATSHGTNNERKLLIEGFPKQTIIFESLDEAGFTFIYQCFTPTTIFYNLNRK  
 LKYLTFHDYGLQFKCKCKEGNLPNYVVVQENRWYDILLNPANDDHPSHDVSQGLVK

EYVALRSSQWNEILLFIITYDEHGGFYDHPVTPLDGVNPDGILGPPVNEFENRLG  
 VVPTFFISPIEPGTVLHSGNPTLMSQTEHSSIPATVKKIFLAKDLTKRDSWAGT  
 FESVITRNSPDQCPETLSNPKMRGTVAKENAELSDFOEELVIAAAGLKGDTKNEEL  
 LYKLCRTYVSDASKYVTKAFDFKFEESKKAREGGDENDIVFCVDDDDDDHNVKPPPP  
 SOSPSHATPWSN"  
 complement(14488..14563)  
 /note="exon predicted by xgrail, quality  
 excellent\_shadowexon"  
 14893..14935  
 /note="exon predicted by xgrail, quality marginal"  
 join(<15141..15586,15675..15872,16258..16667,16780..16950,  
 17036..>17427)  
 Query Match 4.4%; Score 64.8; DB 7; Length 82289;  
 Best Local Similarity 62.2%; Pred. No. 8.5e-05;  
 Matches 102; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
 QY 986 caccacaatccatgcaccaccacggtaacgtctctctacgtgaggtctctcgaacagat 1045  
 Db 12808 CCTAAACGATGATCATCCATCATGACGTCTCCGAAGCCCAAGCTCGTGAAGAAGT 12867  
 QY 1046 ctatgatgccttcgcacggccctcagtgaggacagaccctgctctctctacacga 1105  
 Db 12868 TTACGAGGCTTCGATCGAGTCTCAATGGAATGAGATTTTATTATTAACATATCA 12927  
 QY 1106 cagacacgggtgcttcacgacctgctctccctccctctcgcgcg 1149  
 Db 12928 CGAACACGGTGGATCTTAGCACCATGTCCTACTCCCTCGACG 12971  
 RESULT 2  
 ATAC009327 89934 bp DNA PLN 21-OCT-1999  
 LOCUS Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence,  
 complete sequence.  
 ACCESSION AC009327  
 VERSION AC009327.6 GI:6091751  
 KEYWORDS HTC.  
 SOURCE Arabidopsis thaliana  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 89934)  
 Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.,  
 Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E.,  
 Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.  
 Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence  
 Unpublished  
 2 (bases 1 to 89934)  
 Lin, X. and Kaul, S.  
 Direct Submission  
 Submitted (16-AUG-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org  
 3 (bases 1 to 89934)  
 Lin, X.  
 Direct Submission  
 Submitted (21-OCT-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 On Oct 21, 1999 this sequence version replaced gi:5868934.  
 Address all correspondence to:  
 Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr.  
 Rockville, MD 20850, USA  
 e-mail: xlin@tigr.org  
 BAC clone T12J13 is from Arabidopsis chromosome III and is near the  
 molecular marker mlr2.  
 The orientation of the sequence is from SP6 to T7 end of the BAC  
 clone.  
 Genes were identified by a combination of three methods: Gene  
 prediction programs including GRAIL (available by anonymous ftp  
 from arthur.epm.ornl.gov), Genefinder (Phil Green, University of



```

/ gene="T12J13.4"
/ note="similar to phosphoribosylanthranilate transferase
GB:CAA16616 [Arabidopsis thaliana],
phosphoribosylanthranilate transferase GB:BAAL3032 [Pisum
sativum]"
/ complement(15153..18206)
/ gene="T12J13.4"
/ codon_start=1
/ product="putative phosphoribosylanthranilate transferase"
/ protein_id="AAF03465.1"
/ db_xref="GI:6091755"
/ translation="MADNVLRLKLIIVEICSNRLMPKDGQGTASAYAIYDFDQRRRTK
TFRDLNPFOWDEKLEEFVHDVATMGEEILEINLCNDKTKGRSTFLGKVKIAGSAFAS

Query Match
Best Local Similarity 4.4% Score 64.8; DB 7; Length 89934;
Matches 102; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 986 caccacatccatgcacccacccggtgaacgtctctacgtgaggtctctctcgaacagat 1045
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78527 CGCTAACGATGATCATCATCATGACGCTCTCCGAAGGCCAAAGCTCGTGAAGAAGT 78586

Qy 1046 ctatgatccattcgcaggccctcagtgaggacagaccctgctctctcattaccatcaga 1105
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78587 TTACGAGCCTTGAGATCGAGTCCTCATGATGAGATTTATTCATTATACATATGA 78646

Qy 1106 cgagaccggtggtctacgaccatgctccctccctccctccgcg 1149
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78647 CGAACACGGTGGATTCTAGCACCATGTCCCTACTCCCTCGACG 78690

RESULT 3
LOCUS ATTP8P19/c
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone T8P19.
ACCESSION AL133315
VERSION AL133315.1 GI:6523080
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 96679)
Cholsne, N., Robert, C., Brottier, P., Wincker, P., Catolico, L.,
Artiguenave, F., Saurin, W., Weissenbach, J., Mewes, H.W., Lemcke, K.,
Mayer, K.F.X., Quetier, F. and Salanoubat, M.
Unpublished
2 (bases 1 to 96679)
EU Arabidopsis sequencing project.
Direct Submission
Biochemie, Am Klopfersplitz 18a, D-82152 Martinsried, FRG. E-mail:
lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de
Coordinator: Marcel Salanoubat and Francis Quetier, Grouperment
d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
Gaston Cremieux, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
Location/Qualifiers
1. 96679
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
36..963
/ gene="T8P19.10"
36..86
/ gene="T8P19.10"
/ number=1

CDS
/ gene="T12J13.4"
/ note="similar to phosphoribosylanthranilate transferase
GB:CAA16616 [Arabidopsis thaliana],
phosphoribosylanthranilate transferase GB:BAAL3032 [Pisum
sativum]"
/ complement(15153..18206)
/ gene="T12J13.4"
/ codon_start=1
/ product="putative phosphoribosylanthranilate transferase"
/ protein_id="AAF03465.1"
/ db_xref="GI:6091755"
/ translation="MADNVLRLKLIIVEICSNRLMPKDGQGTASAYAIYDFDQRRRTK
TFRDLNPFOWDEKLEEFVHDVATMGEEILEINLCNDKTKGRSTFLGKVKIAGSAFAS

CDS
join(<36..86,324..475,546..963)
/ gene="T8P19.10"
/ codon_start=1
/ product="hypothetical protein (fragment)"
/ protein_id="CAB62339.1"
/ db_xref="GI:6523081"
/ translation="KTSVKELDEALALTERQHINKLMKEAKARGEGYKIDKLRNTEM
DEYDFLHRSLSREERALLRDLISQAALGLPLEPGRYKPGSFQKQYDPTSAFYQ
DYKGEKNSLSEIKQERMDKHNKSTVGKGNVWDMSYDIDAIKOTIEKRKSGSTLASOE
BETSESEEREDDDDFDDYLSLSESSICYSBQQPLVNGTQVLTD"
87..323
/ gene="T8P19.10"
/ number=1
324..475
/ gene="T8P19.10"
/ number=2
476..545
/ gene="T8P19.10"
/ number=2
546..963
/ gene="T8P19.10"
/ number=3
complement(3123..4007)
/ number=1
3123..4007
/ gene="T8P19.20"
complement(3123..4007)
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="CAB62340.1"
/ db_xref="GI:6523082"
/ translation="MDGRSGCCIAIYAGSGPYDLSKADRIMLRFRPIAPKASPGV
NPVSGSDSGSSDVSPRSASRKRKHQKENGNAKCTRRKTSKPVVHGSANAV
TSLLLPEKPIDLKAQVEKQKROKPLWLSFKDGGCGMLTPAYQTPEIVQVTVISSMPTV
ERYTADWDGVLGRSDEBRKNLVRDTCPGFISDGSRVMTNDAYRKWARDIIPVE
EGAPETSDSFHIVRLVMRPMITSPGTCRMKLYTCQNRGSGVTPVCDVWRM
DVGGEFARLDVKAALCL"
10146..10231
/ note="Tyr-tRNA: anticodon GTA"
10659..12179
/ gene="T8P19.30"
complement(10659..12179)
/ number=1
complement(10659..12179)
/ note="strong similarity to cytochrome P450 CYP94A1 -
Vicia sativa, PIR2:T08014; contains EST gb:H76099"
/ codon_start=1
/ product="cytochrome P450-like protein"
/ protein_id="CAB62341.1"
/ db_xref="GI:6523083"
/ translation="MAFLISFLILAFILITIFFLSSSTKKKVOENTTYGPPSYPLIGS
ILSFKNRRLQWYTELLRSLSPSTILVPLGNRRITITNPLNVEYLKTNFNFPP
KGPFFDLGLDGLGGIFNVDSHSSORRKLASHFEFTRSLRSFAFEVLKDEENRIV
PVLSTAADVTGTVLDQVLKRAFDVCKVSLGMDPCDLDRVNPVLPVAFDPTAAEI
SARRATEPIYAVMTKRVLVNVSERKLEAIRTHVILVSEIVRAKKSLSEIKLEEDV
ODLSRFLAAGHGAVRDMVISFINAGRDTTSAAMTWLFWLLTENDVDVERKILAEVD
PLVSLGLGEDLKEMATKACLEAMRLYPVSWDSKHAANDVLPDGRVYKRGDKVT
YFPTGMGRMETLWGTDSSEFNRFDSFGSTRVYLKPISPKFPVFOAGPRVCVK
EMAFPMQKYVSVLSRSEIIVPNKORPVFVPLTAAHAGLAKVKIKRRSHILNVV"
14300..14811
14300..19503
/ note="copla-like LTR retrotransposon"
23114..25147
/ gene="T8P19.40"
join(23114..23318,23701..23873,23992..24147,24241..24516,
24595..24758,24847..25147)
/ gene="T8P19.40"
/ note="EST GB:T42970 spans first intron, EST GB:N96487
spans third intron; similarity to probable transcription
regulator protein - Sorghum bicolor, PIR2:T03446; Contains
prokaryotic membrane lipoprotein lipid attachment site
AA347-357"
/ codon_start=1

LTR
misc_feature
gene
CDS

```



```

QY 1073 gtggacaagaccgtcttcattacacgagagaccggtggtcttctacgacctgt 1132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1291 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1133 cctccctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcacac 1175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1351 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1393
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
CEP56H9/c 28291 bp DNA INV 14-DEC-1999
LOCUS Caenorhabditis elegans cosmid F56H9, complete sequence.
DEFINITION Z74473.1 GI:1405511
ACCESSION HTG; Guanine nucleotide-binding protein.
VERSION Caenorhabditis elegans.
KEYWORDS Caenorhabditis elegans.
SOURCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
ORGANISM Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 28291)
REFERENCE
AUTHORS none
TITLE Genome sequence of the nematode C. elegans: a platform for
JOURNAL Investigating biology. The C. elegans Sequencing Consortium
MEDLINE Science 282 (5396), 2012-2018 (1998)
REMARK 99069613
The C. elegans Sequencing Consortium.
Erratum: [[published errata appear in Science 1999 Jan
1:283(5398):35 and 1999 Mar 26:283(5410):2103 and 1999 Sep
3:285(5433):1493]]
2 (bases 1 to 28291)
Burton, J.
Direct Submission
Submitted (20-JUN-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1HQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematoe.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormacces&sequence=subject=F56H9
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we arrange for a small
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone F56H9.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone F56H9 is at 1 in this sequence. The true
right end of clone F56H9 is at 12141 in
sequence Z74476.
The true left end of clone W05B10 is at 28193 in this sequence. The
true right end of clone ZC116 is at 18114 in this sequence. The
start of this sequence (1..101) overlaps with the end of sequence
Z74046.
The end of this sequence (28195..28291) overlaps with the start of
sequence Z74476.
Location/Qualifiers
1..28291
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="v"
/clone="F56H9"
complement(join(4691..4961,5062..5143,5193..5622,
5847..7276,7677..7818,8319..8491,8538..8619,8670..9372))
/feature
source
gene

```





```

CDS
join(21125...21361,22361...22675)
/genes="Y46C8AL.3"
/notes="contains similarity to TR:076186; coded for by C.
elegans cDNA yk28g1.3"
/codon_start=1

Query Match
Best Local Similarity 45.7%; Score 52; DB 31; Length 30960;
Matches 181; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

QY 734 caagaatgacgagcttctcgaatgattgtatctctagcaagtcctatcttcagagccgc 793
Db 9675 CAACTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACTCCG 9616
QY 794 caagcgaagggcggtctcgtcctcaactacatgagcgaagcgaattgaaaccca 853
Db 9615 CTACCTGAAGTCTACTCCAACTACTCCGACTACTCCGACTACTCCGACTACTCC 9556
QY 854 ttctcttttctcaactacgtaacacagacctcccggtcccaagtggtgcccgttgaaaa 913
Db 9555 CTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACTCCG 9496
QY 914 ctcttccaagcgcctacgtcgtggtctcctccctaaattctcttattacacccctcgt 973
Db 9495 CTACTCCAACTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACTCCGACT 9436
QY 974 ctgcggcaccacacaaactcctcatgacacccacccgtaacggtctctcagtgaggtctt 1033
Db 9435 CTCCGACTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACT 9376
QY 1034 cgtcaagcagatctatgatgccattccgagggccctcagtgaggaagacccctcctt 1093
Db 9375 CTCAACTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACTCCGACTACT 9316
QY 1094 cattactcagcagagacgggtggtcttcaagacca 1129
Db 9315 CGACTACCATGAAGTCTACCAACTACTCCGACTACTCCGACTACTCCGACTACT 9280

RESULT 7
AF107252
LOCUS Burkholderia pseudomallei non-hemolytic phospholipase C gene,
DEFINITION complete cds.
ACCESSION AF107252
VERSION AF107252.1 GI:6572520
SOURCE Burkholderia pseudomallei.
ORGANISM Burkholderia pseudomallei.
REFERENCE Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
AUTHORS Burkholderia; pseudomallei group.
1 (bases 1 to 2262)
Korbsrisate,S., Suwanasai,N., Leelaporn,A., Ezaki,T., Kawamura,Y.
and Sarasombath,S.
TITLE Cloning and characterization of a nonhemolytic phospholipase C gene
JOURNAL from Burkholderia pseudomallei
MEDLINE J. Clin. Microbiol. 37 (11), 3742-3745 (1999)
PUBMED 99454884
REFERENCE 10523590
AUTHORS 2 (bases 1 to 2262)
Korbsrisate,S., Suwanasai,N., Leelaporn,A., Kawamura,Y., Ezaki,T.
and Sarasombath,S.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-1998) Department of Immunology, Fac. of Medicine
Siriraj Hospital, Mahidol University, Prannok Rd., Bangkok 10700,
Thailand
FEATURES
source Location/Qualifiers
1..2262
/organism="Burkholderia pseudomallei"
/strain="SPT1"
/db_xref="taxon:28450"
/clone="pSN-1"
111..116
RBS

```

```

CDS
124..2226
/transl_table=1
/product="non-hemolytic phospholipase C"
/protein_id="AAFI7299.1"
/db_xref="GI:6572521"
/translation="MTNORRDFLRLAGTAGAAGALQLFPVIREALAIPIANRRGTGI
RDVEHIVLMQENRSEDFHFKLGRVGRGDEPRPLTLQNGRPFHQHVLGPAELLFP
HPDASNLGHWQFQDLPHGQDTHGAMNKGRYDRWIAKGTMTTAYLERGDIPTFYQLA
DAFTICDAHCSTIPSDPNRYMTWTQYVNDGAGGPGVNGEAGYGSWYPTLEQ
AGVSKIIYQDITGDLDAAGSWGTONPYIGNYDALLYFNQYRNAQPGSPYLDKALT
GTNSAGGTLFDVLDQDVKNGTLPQVSWICAPAYSEHPNPNYAGWAVEQVLKRT
SPDWNKSTALITIDYDENDGFHDVAPFAPQSRDNGLSTVSTAGELFPDAAHWAGP
YGLGPRVPLVSPMTKGGWVCSQTDHTSLQFTEARNDRYPVRAPNVTPWRAVC
GDLTSPAFNLSPPGDSWLPDTSYATPPDRNRHPSYVPVPAQSMFKPQKAGRAARA
LPYELFVLGRIDQSTGKFKLFAKTGAGAAQVYAGNRVDGPMWATVBARLRSDEN
STALTSYDLTVYGPNGFLCPRGSTAAALGNANPEVIYDYDVANGNITLRLSNRG
RATVRLTVANAYKSPRVYELAPGQRINDYIDLHSDSHSYDTUTVSDGAPNGFLRREA
GHVETGRPSTSDPLIATV"
BASE COUNT 388 a 794 c 752 g 328 t
ORIGIN

Query Match
Best Local Similarity 3.4%; Score 50.6; DB 1; Length 2262;
Matches 111; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

QY 974 ctgcggcaccacacaaactcctacgacccaccacgggtacgctctcctcagtgaggtctt 1033
Db 1038 CTGCGCGCGGAAGCGTATTCTCGAGGACCCCGAAGTGGCGGCGGTGGTA 1097
QY 1034 cgtcaagcagatctatgatgccattccgagggccctca---gtgggacaagacctgct 1090
Db 1098 CGTCGAGCAGGTGCTGAAGGCGCTGACGTCGATCCGAGCGTGTGGACCAAAACCGCGCT 1157
QY 1091 ctctcattcctcagacagacgggtggtcttctcagaccatgctctccctctcgcctt 1150
Db 1158 CTTCATCAGTACGAGCAGGACGAGCGGCTTCTTCGATCAGCTCGCGCGCGCTCGCGC 1217
QY 1151 cgcgcccgagacaac 1163
Db 1218 GCAATCGCGCGAC 1230

RESULT 8
AC020448/c
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC020448
VERSION AC020448.1 GI:6664449
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 2426)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10213199 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source Location/Qualifiers
1..2426
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
594 a 417 c 670 g 745 t
BASE COUNT

```







```

JOURNAL      structural proteins of the Thermoproteus tenax virus TTV1
MEDLINE      Mol. Gen. Genet. 217 (1), 105-110 (1989)
REFERENCE    89364672
AUTHORS      Neumann, H. and Zillig, W.
TITLE        3 (bases 12541 to 13242)
             The TTV1-encoded viral protein TPX: primary structure of the gene
             and the protein
JOURNAL      Nucleic Acids Res. 18 (1), 195 (1990)
MEDLINE      90174928
FEATURES     Location/Qualifiers
             1..13669
             /organism="Thermoproteus tenax virus 1"
             /strain="KRA1"
             /db_xref="taxon:10479"
             292..489
             /note="URF (65 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32969.1"
             /db_xref="GI:62154"
             /db_xref="SWISS-PROT:P19276"
             /translation="MWDPDIEVESIPPEIMEEVEGEYIDIFQLTEIAEMERDVIIVEE
             QURDEAKQFEYEDFWDYVT"
             486..869
             /note="URF (127 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32970.1"
             /db_xref="GI:62155"
             /db_xref="SWISS-PROT:P19277"
             /translation="MTRVNYIRNLIRGEIEIKGVSSDRIPMIEIEPGHNDPERMCV
             MFMNEYSHIPQADNLDEVDYDCKILTSKITTYINVAIFGKNGKLGKLSMLMLVTRD
             KSKREKVCLELGKIKAYNWTCKPQ"
             838..1011
             /note="URF (57 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32971.1"
             /db_xref="GI:62156"
             /db_xref="SWISS-PROT:P19278"
             /translation="MHIIGPVFSRVTFLDFVNYLQSGLYETLIRDKYIEYTLKMK
             YATLRLYGKYVHT"
             989..1180
             /note="URF (63 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32972.1"
             /db_xref="GI:62157"
             /db_xref="SWISS-PROT:P19279"
             /translation="MASTFTHDIVMECYEEALAVMRCEVAYDRVLTWKMEPSLPLLR
             LDNEIIPYMKVYKYYQL"
             1289..1534
             /note="URF (81 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32973.1"
             /db_xref="GI:62158"
             /db_xref="SWISS-PROT:P19280"
             /translation="MAELNKAIKLTIKYKIAMNYPYTADIAEKLTIELGLIRQEKWLKAI
             NDKRPKDLIDKRNKNKTEKSENTGENHDNDQQR"
             complement(1467..1688)
             /note="URF (73 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32975.1"
             /db_xref="GI:62160"
             /db_xref="SWISS-PROT:P19281"
             /translation="MRHAPSVLSRLFTIVDFVWDFSMIRPCTICTWTYSISCTGYILL
             ILIYRFSFGLDYDHSLLYSILFYC"
             1509..1850
             /note="coat protein TP1 (AA 1 -113)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32974.1"
             /db_xref="GI:62159"
             /db_xref="SWISS-PROT:P19270"
             /translation="MIIITKNREYTTIDENKEINSISAADTICPYRAYRSYHRKIPIK
             IDCKEARKTAGSMTHYVILRQLKEQCCEKTEKIVDPKLFRADAICDGDALLRLRQTS
             IGTGMRLISGN"
             complement(1771..2037)
             /note="URF (88 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32977.1"
             /db_xref="GI:62162"
             /db_xref="SWISS-PROT:P19283"
             /translation="MLHHPILFLSSTSPFFIRLIRFLTYSVIFTRSSSSICIVWLMP
             LLMYIRCPILSAPASAIYTFNCHLLSACQSRKLFVLLSTMPPHHI"
             complement(1889..2065)
             /note="URF (58 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32978.1"
             /db_xref="GI:62163"
             /db_xref="SWISS-PROT:P19284"
             /translation="MITFSFKFAYVTPTSSIPLLHSFISYSSVOISNLSNFSYLLF
             VHLYCRPYAIVNVY"
             1992..2162
             /note="URF (56 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32976.1"
             /db_xref="GI:62161"
             /db_xref="SWISS-PROT:P19282"
             /translation="MEKLMRRGIEEYGYVYANKLVNVIINCYEGERSHRDTGNNIHIHC
             SSAKPYSLLCFT"
             2074..2286
             /note="URF (70 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32979.1"
             /db_xref="GI:62164"
             /db_xref="SWISS-PROT:P19285"
             /translation="MKAKKEVLEIITIIYVVPVSLIVVYVSHNVYVAIVTTLNL
             ARKTYTKKIEKMKKKEVALEFFS"
             complement(2138..2389)
             /note="URF (83 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32981.1"
             /db_xref="GI:62166"
             /db_xref="SWISS-PROT:P19287"
             /translation="MLPLMAISSLGIGTSSSPRYTILSSLIIVYTLMKRTRAPFL
             LSSSFQSFSLYKSSLSLNSMLPLLLHPIYVYKHNKL"
             2283..2582
             /note="URF (99 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32980.1"
             /db_xref="GI:62165"
             /db_xref="SWISS-PROT:P19286"
             /translation="MRVYTRDDRIYVLGEYELVPMPKLEDDIAAIKGSIIAYVLNSN
             YSSFGALKDRTKEMELIENLAYELHQNDRTPTLTRYYIYLLMKVYGMHGSILSQF"
             2672..3076
             /note="URF (134 AA)"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA32982.1"
             /db_xref="GI:62167"
             /db_xref="SWISS-PROT:P19288"
             /translation="MKFINLLMLMSVVVGRIAGANELEMRYQAGIEVKRQDVQVKR
             VERKISYRAPRIEDAIINLYEDKISGKPYIGPLTGLVRKAIKIAYRDGVDSAMNI
             FKSYDAIONEIRKIIIEGLKMWKIRSEQR"
             3073..3498
             /note="URF (141 AA)"
             /codon_start=1

```

```

/transl_table=11
/protein_id="CAA32983.1"
/db_xref="GI:62168"
/db_xref="SWISS-PROT:P19289"
/translation="MSPPIISNYLEGIRDNKYLQRIEWEIQVAGSGFRAIPITERVMWG
EASYVIDLANKIKALPELSNYKTRPIGYALAAARYVKKHNFMAEIIYDRFESALVRS
GLSRDMARRVREYAKMIGIGGGGGGGGGGGGGVSLUP"
complement(3226..3417)
CDS
/note="URF (63 AA)"
/codon_start=1
/transl_table=11
/protein_id="CAA32985.1"
/db_xref="GI:62170"
/db_xref="SWISS-PROT:P19291"
/translation="MLSNPSHRSQATDKRLKSIKYLGHAEVVVFNVRGQRI
SNELSLNIAQFGQSLLVIG"
3263..3430
CDS
/note="URF (55 AA)"
/codon_start=1
/transl_table=11
/protein_id="CAA32984.1"
/db_xref="GI:62169"
/db_xref="SWISS-PROT:P19290"
/translation="MLRLSLDMKWRPRDTLNTTTSAMPRTYTLIDLSLRLSVADCREI
WRGLSWMQK"
3509..3940
CDS
/note="URF (143 AA)"
/codon_start=1
/transl_table=11
/protein_id="CAA32986.1"
/db_xref="GI:62171"
/db_xref="SWISS-PROT:P19292"
/translation="WKYFAIGFIFSPSEFLYNDYKIRTIYTVRVGGIDREVRVG
IAHMEGEVRIIEGIPITVRGDKTSASIIYSGNWNISMYNLSLRFIWAYNAKATV
IRPTGRARGEGCLRVLDEIDAGVYSGRLKPTIVQLREE"
4041..4460
CDS

Query Match      3.3%; Score 49.4; DB 80; Length 13669;
Best Local Similarity 45.6%; Pred. No. 0.33;
Matches 173; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 773 caagtcctctcagcgcgaagagaggggtgtctcgtcgaactacgtagtggcag 832
Db 11947 CGACCTACGCCGACCTCCACATCCACACCTACTCCGACTCCGACCTACTCTCAACAC 12006

QY 833 caacgagaaattcgaacggattctcttcttcaactacgtcaacagacactcccggtc 892
Db 12007 CAACGCCAATCCGACACCGACTCCCAAGCCCAAGCCCTACTCTACACCAACCCCAACAC 12066

QY 893 caacgtggtgcccgttgaaactcttccaagacgctcactcgtgtctccctcaatt 952
Db 12067 CTACCCCTACTCTACACCTACGCCGACTCCGACTCCGACACCCGACTCTCTACGCCCTACTC 12126

QY 953 ctcttactaataacccctctgtcgtcgcacaaacaaactcattgcacccacccggttaa 1012
Db 12127 CGACTCTTACCCCAACTCTCTACTCTCCACTCCACCCGCTACTCCGACTCCACACCGACTC 12186

QY 1013 cgtctctacggtgaggtcttcgtcaacgagatctatgatgcattgccaggagccctca 1072
Db 12187 CGACACCTACTCCGACGCTACTCCGACTCCACGCGGCTACTCTTACACCAACCCCTTAC 12246

QY 1073 gtggagaaagacctgtcttcttactacacgagacgggtggtcttctacgacctatg 1132
Db 12247 CTACACCAACTCCACCCCTACTCTCGACTCCGACGCGCAACTCCGACTCCCAACGCGC 12306

QY 1133 cctctccctctcgtccgtc 1151
Db 12307 CTACACCAACTCCGACGCTC 12325

```

RESULT 11  
F10K1  
LOCUS

DEFINITION Sequence of BAC F10K1 from Arabidopsis thaliana chromosome 1, complete sequence.

ACCESSION AC067971

VERSION AC067971.4

KEYWORDS GI:8569088

SOURCE HTG.

ORGANISM Arabidopsis thaliana

REFERENCE

AUTHORS

Li, S.X., Chan, A., Yu, G., Etgu, P., Lee, J.M., Lenz, C., Pham, P., Sakano, H., Toriumi, M., Chung, M., Goldsmith, A., Liu, A., Liu, A., Vayberg, M., Altaf, H., Brooks, S., Buehler, E., Chao, Q., Conn, L., Conway, A., Hansen, N., Johnson-Hopson, C., Khan, S., Kim, C., Lam, B., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R., Federspiel, N.A., and Theologis, A.

TITLE The sequence of BAC F10K1 from Arabidopsis thaliana chromosome 1 unpublished

REFERENCE

AUTHORS

2 (bases 1 to 104533)

JOURNAL Theologis, A.

TITLE Direct Submission

JOURNAL Submitted (27-APR-2000) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

REFERENCE

AUTHORS

3 (bases 1 to 104533)

JOURNAL Theologis, A.

TITLE Direct Submission

JOURNAL Submitted (16-JUN-2000) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

REFERENCE

AUTHORS

4 (bases 1 to 104533)

JOURNAL Theologis, A.

TITLE Direct Submission

JOURNAL Submitted (07-JUL-2000) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

REFERENCE

AUTHORS

5 (bases 1 to 104533)

JOURNAL Theologis, A.

TITLE Direct Submission

JOURNAL Submitted (04-OCT-2000) Plant Gene Expression Center, 800 Buchanan St., Albany, CA 94710, USA

COMMENT On Jun 16, 2000 this sequence version replaced gi:8099274. The sequence is of BAC F10K1 from Arabidopsis thaliana chromosome 1. The sequence does not represent the sequence of the entire insert of this clone. It is shorter by 71175 bp because we submit only the unique sequence of the clone. However, in order to facilitate the joining of overlapping clones in the future for creation of larger contigs, we provide small overlaps (200 bp) between overlapping submitted clones. The 5' end of this sequence overlaps by 200 bp the 3' end of the sequence of the clone F22G5. The 3' end of this sequence overlaps by 200 bp the 5' end of the clone F4H5.

FEATURES

source

1..104533

/organism="Arabidopsis thaliana"

/db\_xref="taxon:3702"

complement(<1..924)

/gene="F10K1.1"

complement(<1..924)

/gene="F10K1.1"

/note="Strong similarity to an unknown protein At2g29670 gi|3582340 from Arabidopsis thaliana BAC F27A16 gb|AC005496. EST gb|R90693 comes from this gene. this gene may be cut off."

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AAF82193.1"

/db\_xref="GI:8954019"

/translation="MGVKVATSTTHQVAHPVIVHSSLSQTLASSAVSSPSRRRII GNDGRSLSCRSVMQSORLNPSPGTSSTNLRKSKSELWETKRPKTQIIRAFSA NYDPSEEEFAKKMOELTKQVSNNEEDNESRTRDDTRMDIVGSHNFKFSME PWPMEVQMSNIERKANSVDLPLSLRIKKRQMEGVNLNVQGSACCSYTKAFSSW FMRRELASFLLHMRKELLFEDDQGLHVRKEMQSFVNLQVQVFSATPTLMVSMVL LANFTVYSIESNSALAAVSPPTLTLSFSFTT"

6532..9581

gene



```

QY 1239 acctgggtatctcccttacagtaagaagg 1270
||||| | ||||| | | | |||||
Db 17713 ACCCTTCTGATCTCTCTGGATTGAGAAGG 17744

RESULT 12
TTV1VP 4034 bp DNA VRL 10-FEB-1999
LOCUS Thermoproteus tenax virus 1 viral protein TPX-VT3 exhibiting
DEFINITION genomic variation.
ACCESSION X14717.1 GI:62191
VERSION X14717.1 GI:62191
KEYWORDS unidentified reading frame.
SOURCE Thermoproteus tenax virus 1.
ORGANISM Thermoproteus tenax virus 1.
VIRUSES; dsDNA viruses, no RNA stage; Lipothirixviridae;
Lipothirixvirus.
1 (bases 1 to 4034)
REFERENCE
AUTHORS Neumann, H.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-1989) Neumann H., Max-Planck-Institut fuer
Biochemie, Am Kloeperspitze, 8033 Martinsried, F R G
REFERENCE
2 (bases 1 to 4034)
AUTHORS Neumann, H. and Zillig, W.
TITLE Nucleotide sequence of the viral protein TPX of the TTV1 variant
VT3
JOURNAL Nucleic Acids Res. 18 (8), 2171 (1990)
MEDLINE 90245666
COMMENT See <X14855> for complete genomic sequence.
Many of the URF's begin with GTG instead of ATG.
FEATURES
source Location/Qualifiers
1. 4034
/organism="Thermoproteus tenax virus 1"
/strain="KRAL 10/12"
/db_xref="taxon:10479"
132..1536
/notes="viral protein TPX-VT3 (AA 1-474)"
/codon_start=1
/transl_table=11
/protein_id="CAA32838.1"
/db_xref="GI:584349"
/db_xref="SWISS-PROT:P19275"
/translation="MSYQITNNNTSNQVNIIVADPNLASLGENVVFSDSQPLSL
YVVDGVVJIKRSLPSGQSIITASSGPNIDPTIALYNNNGSSYNLTGLVSPVVS
IVDQCGYAIASAYASGDFLVAQPTGTFPSSRLVVDWRAFTPSLDAVGLRIYADT
NDMPGVVRKYVNGAQNVSIEQIKSGHYSVNEIDISQFAAFTDPLVMVLSINGSTANV
KYVKGSGNIGTVSGNYSTIEYGNPSMAGYGVDRKHYANFVLPYDPQVTVPISSP
SPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
complement(533..811)
/notes="unidentified reading frame (AA 1-52)"
/codon_start=1
/transl_table=11
/protein_id="CAA32839.1"
/db_xref="GI:62193"
/db_xref="SWISS-PROT:P19306"
/translation="MFEPCLYTFEAVLPFDIXYINGSVNAANDISISFTTEYVPD
IFCSMLTP"
complement(890..1243)
/notes="unidentified reading frame (AA 1-117)"
/codon_start=1
/transl_table=11
/protein_id="CAA32840.1"
/db_xref="GI:584350"
/db_xref="SPTREMBL:O88913"
/translation="MISYGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
VGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
VTCGSGSGNITKLA"
1991..2311
/notes="unidentified reading frame (AA 1-106)"

CDS
Query Match 3.2%; Score 47.8; DB 80; Length 4034;
Best Local Similarity 48.1%; Pred. No. 0.82;
Matches 136; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
QY 757 actatggtatctctagcaagtcattcttcgagcccaagagagggcgctgcctggc 816
||||| | ||||| | | | |||||
Db 942 ACTCTATATCTTCAACATCTCTTACTCTACACCGACACCGACGCTACACCC 1001
||||| | ||||| | | | |||||
QY 817 tcaactacgatggcaccacgaggaattcgaacggattctctcttcttcaactacgta 876
||||| | ||||| | | | |||||
Db 1002 ACACCAACGCCAATCCGAGCGCTACTCTTACCCCTACTCCGACACCTACCGACTCC 1061
QY 877 accagacctcccggtccaacgtygtgcccgttgaaacttcttccaagcgcctaactcg 936
||||| | ||||| | | | |||||
Db 1062 ACTCCACACCTACTCCGACTCCGACACCTACTCCAAACACCAACGCAACTCCGACACCG 1121
QY 937 gtgtcctctcttaattcttctacattacccctctctgtcgccaccacacccaactcca 996
||||| | ||||| | | | |||||
Db 1122 ACTCCAAGCCACGCGCTACTCCGACTCCGACTCCGACTCCGACTCCGACTCCGACTCC 1181
QY 997 tgcaccaccacggtaacgtctctctactacggtgaggtcttctgtaaa 1039
||||| | ||||| | | | |||||
Db 1182 ACTCGGACTCCAAACACCGACTCCGACACCAACGCGCTACGCGCTA 1224

RESULT 13
NPORPT2 749 bp DNA VRL 03-FEB-1999
LOCUS Multiplicapoid nuclear polyhedrosis virus of Orgyia
DEFINITION pseudotsugata (OpNPV), GC-rich repeated DNA from the SstI I

```

```

fragment.
ACCESSION D13755 D00297
VERSION D13755.1 GI:222216
KEYWORDS repetitive GC-rich region.
SOURCE Multicapsid nuclear polyhedrosis virus of Orgyia pseudotsugata DNA,
clone cosmid 58.
ORGANISM Orgyia pseudotsugata nuclear polyhedrosis virus
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
REFERENCE 1 (bases 1 to 749)
AUTHORS Chen,D.D., Nesson,M.H., Rohrmann,G.F. and Beaudreau,G.S.
TITLE The genome of the multicapsid baculovirus of Orgyia pseudotsugata:
restriction map and analysis of two sets of GC-rich repeated
sequences.
J. Gen. Virol. 69 (Pt 6), 1375-1381 (1988)
JOURNAL 88258472
MEDLINE
COMMENT Analysis of the region from the SstI I fragment that was
crosshybridized with the repeated DNA downstream of the polyhedrin
gene revealed a sequence made up predominantly of repeats of the
sequence GCGC or the closely related variant GGTG, and a
dinucleotide which results in amplification of groups of six
nucleotides.
FEATURES
source Location/Qualifiers
1. .749
/organism="Orgyia pseudotsugata nuclear polyhedrosis
virus"
BASE COUNT 111 a 163 c 289 g 186 t
ORIGIN 5 up upstream of PstI site.

Query Match 3.1%; Score 46; DB 80; Length 749;
Best Local Similarity 44.3%; Pred. No. 2.3;
Matches 187; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

Qy 756 aactatggtatcttagcaagtcattcttcgagcgccgaagagagggcggtctctg 815
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 630 AACAAAGCCACCTTTTGGCTTCGCGGACGCGCGCAACGCCAACGCCGCGGACGCC 571

Qy 816 ctcaactacgtagcaccacgaagattcgaacggtattctcttcttcttctcactc 875
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 570 AAGCCATCGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCC 511

Qy 876 aaccagacacctccgggtccaaagtgtgcccgttgaaacttcttccagagcgctactc 935
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 510 ATCGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 451

Qy 936 ggltgctcccttaattctcttaccattacacccctctctgctgagcgccacacacacac 995
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 AAGCCATCACCACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGAC 391

Qy 996 atgcacccccacggtaacgtctctctcaggtgaggtctctcagcagatctatgacc 1055
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 GAGGCCATCGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 331

Qy 1056 attcgccaggccctcagtgagcaagaccctctcttcttcttcttcttcttcttctt 1115
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 330 GAGGCCATCACCACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGAC 271

Qy 1116 ggcttctagacacatgctctctctctctctctctctctctctctctctctctctct 1175
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 GAGGCCATCGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 211

Qy 1176 ga 1177
||
Db 210 GA 209

RESULT 14
SC4C6
LOCUS SC4C6 30941 bp DNA BCT 21-JUN-1999
DEFINITION Streptomyces coelicolor cosmid 4C6.
ACCESSION AL079355

```

```

VERSION AL079355.1 GI:5139546
KEYWORDS 3-carboxy-cis,cis-muconate cycloisomerase; 3-oxoadipate
CoA-transferase subunit A; 3-oxoadipate CoA-transferase subunit B;
3-oxoadipate enol-lactone hydrolase; 4-carboxymuconolactone
decarboxylase; acyl-[acyl-carrier protein] desaturase;
beta-ketoadipyl-CoA thiolase; bldA regulation; copper oxidase; DNA
ligase; glutathione-dependent aldehyde dehydrogenase; glycosyl
hydrolase; hydroxylase; laci-family; membrane protein; pcab; pcaf;
pcag; pcab; pcab; pcab; pcab; pcab; pcab; pcab; pcab; pcab; pcab;
3,4-dioxygenase alpha subunit; protocatechuate 3,4-dioxygenase beta
subunit; regulatory protein; secreted protein; transcriptional
regulator.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
REFERENCE 1 (bases 1 to 30941)
AUTHORS Redenbach,M., Kieser,H.M., Denapaltie,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
JOURNAL 97000351
MEDLINE
REFERENCE 2 (bases 1 to 30941)
AUTHORS Seeger,S. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 30941)
AUTHORS James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S.coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.gov/jp/
Jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or att) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid 4C6 Cosmids 4C6 lies between 6G3 and 5F2 on the AseI-A
genomic restriction fragment.
FEATURES
source Location/Qualifiers
1. .30941
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 4C6"

```

```

gene
/ gene="SC4C6.01"
CDS
<1..1506
/ note="SC4C6.01, probable phospholipase C, partial CDS,
len: >501 aa; similar to e.g. SW:PHLN_PSEAE (EMBL:M59304),
p1cN. Pseudomonas aeruginosa non-hemolytic phospholipase C
precursor (692 aa), fasta scores; opt: 1073 z-score:
1149.1 E(): 0, 44.7% identity in 524 aa overlap"
/codon_start=1
/transl_table=11
/label="SC4C6.01"
/product="putative phospholipase C"
/protein_id="CAB45566.1"
/db_xref="GI:5139547"
/translation="NDGAGGPGVLGNDELGYDWTTPERLEAAGVSMKIYQDVGBGLD
AGHWCWIDDAYRGNTGNSLLYFNKYNKAPGDPDYDKARTGDKRGEYFDRIRA
DYKAGKLEISWIAPEAFSEHSNMPYSIWQVLDALTSNPKVWAKTALFITYD
ENDGFEDHVPPLPPKSAAGRSTVDYSLDFEGSAKHREGFYGLGPRVPMLVSPWS
KGGFVCSFEDHTSYIFRMEKRGVHEPNISFWRRAVCGDLTSFDFSRDTPRKLPL
DTDAYEPDRERHPDYRPTPPADPGMPROGRGTPTPLRYAVPHVCGAVDTAAGKFTL
AFASOPRAGAALFVTSNRTDGPVYTTAGKDIAUTNLSAYSCGSVDLTIVHCPAGFL
RVFRGANEVAGPAGPEVTAHRGDDLRLLTLTHTCTGVRLRIANADGRTSRVTVRPG
ATVHTVTGLARSRYDVTVTSETAPAFRLRRFAGHVENGREGTSDFAIITD"
1275..1300
/repeat_unit

/ gene="SC4C6.01"
/ note="repeat 1; 26 bp imperfect inverted repeat"
/complement(1560..2513)
/ gene="SC4C6.02c"
/complement(1560..2513)
/ gene="SC4C6.02c"
/ note="SC4C6.02c, probable integral membrane protein, len:
317 aa; similar to TR:086668 (EMBL:AL031182) S.coelicolor
probable integral membrane protein (299 aa), fasta scores;
opt: 288 z-score: 279.1 E(): 3.2e-08, 32.8% identity in
299 aa overlap and to TR:033234 (EMBL:Z98209)
Mycobacterium tuberculosis hypothetical protein (301 aa)
(28.8% identity in 299 aa overlap). Contains hydrophobic,
possible membrane-spanning regions. Weakly similar to
TR:086668 (EMBL:AL031182) S.coelicolor putative integral
membrane protein (299 aa) (32.8% identity in 299 aa
overlap)"
/codon_start=1
/transl_table=11
/label="SC4C6.02c"
/product="putative integral membrane protein"
/protein_id="CAB45566.1"
/db_xref="GI:5139548"
/translation="MNAALVAVLLSLVSACVAAAQAERLASAGAGALMLASG
AMWVALNSAALLHVVALKYGTLVQPLGALTVAAPVGLARLGRVSAIENRG
TGLTLGLCALLTASGPADDDVLSVPEALTVSGATALLGLVSRPGARGLRHATAS
GIASGVASALTQTVTAVTRDSRGLLSQVIGVAVLVAAPATGGLLLSQTAYRGGLGA
PLAVVTLNPAVAAVIGLSLGLRGRGAGGVLLALITGAALAAWGLQLSRAAPDRPE
TAPGLAGFGLAGTETPTGPAPGPGVVPVPRQPGPHLTRL"
/complement(1605..1630)
/ gene="SC4C6.02c"
/ note="repeat 1; 26 bp imperfect inverted repeat"
/complement(2510..3325)
/ gene="SC4C6.03c"
/complement(2510..3325)
/ gene="SC4C6.03c"
/ note="SC4C6.03c, possible membrane protein, len: 271 aa;
unknown function, probable CDS suggested by positional
base preference, GC frame plot and amino acid composition.
Contains hydrophobic, possible membrane-spanning regions"
/codon_start=1
/transl_table=11
/label="SC4C6.03c"
/product="putative membrane protein"
/protein_id="CAB45567.1"
/db_xref="GI:5139549"
/translation="MNASEPAQAGPDPAPPLVRELLVAGLFLVYKFRRLATGE
TAEAFRNAHHWDMWVRLPDEGAVQSLLHGDTLIHVANTYAAVHFAVTAFLW

```

```

LYLRPAHYLWARRVLAAVTGAALVLLHLLFPLAPPRLAATGLDITARLFGPSVGGPP
RTDLSNQFAFMSPLHFGWALMIAIGIVATRTWRWLLPLVTLVTVGTANIHYW
LIVATVALLGLALAVLRLRPRGRPCALPHRRTRVRRSSAERPELVGAGR"
3410..3459
/ note="stem is 21/22 (95%) matches, 0 gaps; loop is 6 bp"
/ gene="SC4C6.04"
/ gene="SC4C6.04"
/ gene="SC4C6.04"
/ note="SC4C6.04, possible transcriptional regulator, len:
204 aa; similar to many e.g. TR:P72185 (EMBL:U51164),
hemK, Protonibacterium freudenreichii repressor protein
from hemYHXRRL gene cluster (involved in the biosynthetic
pathway from glutamate to protoheme) (259 aa), fasta
scores; opt: 205 z-score: 251.7 E(): 1.1e-06, 28.5%
identity in 200 aa overlap. Contains probable
helix-turn-helix motif at aa 45-66 (Score 1187, +3.23 SD).
Weakly similar to others from S.coelicolor e.g. TR:069987
(EMBL:AL022268) S.coelicolor probable transcriptional
regulator, partial CDS (>111 aa) (42.4% identity in 85 aa
overlap) is similar to the N-terminal half"
/codon_start=1
/transl_table=11
/label="SC4C6.04"
/product="putative transcriptional regulator"
/protein_id="CAB45568.1"
/db_xref="GI:5139550"
/translation="MTSQDADRPAAAGTSRRSKITPEREQEFDVLDQLRSCGYDVA
TMGIAASTRCSTKSTLYRQMKTPQFAAALRSNRVFTDIDTGLSADLROVAAA
GEGAGKDTGLLOALGHVAMDELPKALREALVEPEIAALRAILARGVARGEVPAGHP
ALEYVPAQLFGMLRMPVLEGEQDAAYLVREVAVLPALGLP"
/complement(4284..5327)
/ gene="SC4C6.05c"
/ gene="SC4C6.05c"
/ note="SC4C6.05c, possible membrane protein, len: 347 aa;
similar to TR:088067 (EMBL:AL031541) S.coelicolor possible
membrane protein (319 aa), fasta scores; opt: 474 z-score:
403.9 E(): 3.6e-15, 37.9% identity in 354 aa overlap.
Contains 4x PPXQ repeats starting at aa 46. Contains a
hydrophobic, possible membrane-spanning region. Similar to
TR:088067 (EMBL:AL031541) S.coelicolor possible membrane
protein (319 aa) (37.9% identity in 354 aa overlap)"
/codon_start=1
/transl_table=11
/label="SC4C6.05c"
/product="putative membrane protein"
/protein_id="CAB45569.1"
/db_xref="GI:5139551"
/translation="MTQATPPGWYPDPQKNDGPATERMWDGTANTDRVRPAGCAAAW
APPAQPGQPPVQPPVDTAAGPYVPHPGYPIVQPPSARRRRLRAGIAVAATAVAVL

```

```

Query Match
Best Local Similarity 3.1%; Score 46; DB 29; Length 30941;
Matches 140; Conservative 0; Mismatches 135; Indels 3; Gaps 1;

QY 870 tacgttaaccagacctcccggtcccaacgtgtgtccggttgtaaaactttctccaagcgc 929
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 TACGACAAAGCCCGCACCGCACCGCACCGCACCGCACCGCACCGCACCGCGCTC 300
QY 930 tacctcgtgtgtctctccctaaattcttaccattaaacctctcgtcgcacacacc 989
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 CGTGGCCGAGCTCAACGCGGGCAAGCTCGCGGAGATCTCTCGATAGCGCCCGGAGCA 360
QY 990 aactccatgacccccacgggtaacgtctcctcagtgagggtcttcgtcaagcagatctat 1049
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 TTCTCCGAGCATTCCCACTGGCCGCTCGAAGTACGCGGCTGTATCTCCAGGTCCTG 420
QY 1050 gatgcattcgcaggggccctca--gtgggacaagacctgtcttcattaccatcac 1106
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 GACGCGCTACCTCCACCAACCGAAGGTGTGGCGCAAGCGCCCTGTTTCATCAGTACG 480
QY 1107 gagaccggtggtctctcagacatgtccctccccctct 1144

```



```

Db 481 GAGAACGACGCTTCTTCGACACGCTGGTCCGCCGCGCT 518
||||| || ||||||| ||||||| || || || ||
RESULT 15
LOCUS OPU75930/c 131993 bp DNA circular VRL 06-MAR-1998
DEFINITION Orgyia pseudotsugata nuclear polyhedrosis virus complete genome.
ACCESSION U75930
VERSION U75930.1 GI:2934903
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 131993)
Nucleopolyhedrovirus.
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Rohrmann,G.F.
The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
polyhedrosis virus genome
Virology 229 (2), 381-399 (1997)
72721300
REFERENCE
AUTHORS
2 (bases 1 to 131993)
Rohrmann,G.F.
Direct Submission
Submitted (23-OCT-1996) Oregon State University, Agricultural
Chemistry, Corvallis, OR 97331-7301, USA
3 (bases 1 to 131993)
Rohrmann,G.F.
Direct Submission
Submitted (06-MAR-1998) Oregon State University, Agricultural
Chemistry, Corvallis, OR 97331-7301, USA
Sequence update by submitter
On Mar 6, 1998 this sequence version replaced gi:1911246.
similar to Autographa californica nuclear polyhedrosis virus
complete genome: GenBank Accession Number U22858.
(ACMNPV)
Location/Qualifiers
1. 131993
/organism="Orgyia pseudotsugata nuclear polyhedrosis
virus"
/db_xref="taxon:10450"
complement(123..947)
/note="ORF1; pk-1, ACMNPV ORF10 homolog"
/codon_start=1
/product="protein kinase 1"
/protein_id="AAC59000.1"
/db_xref="GI:1911247"
/translation="MDAALQSLRFRVADCAVLAPKVNGRFGKMDVLHRRPTTSKLEFL
RKTAAHSFSADEINVDHMSDPFVDMYFCYSSPTAWATVMDYVPCDLPFETLQTO
GALNALVNIYROLCDALNDLNATGYIHNQVLENVLYFGARDRVYLCGYLCKRE
HSPVHDGTLSEYSPKIRIRNYARSDPWAVGVLAYKLLTGGRRPERSADEVLDLAS
MRRQQQINDPAAKKNVRLMARDFVCLTRFNFECRSIDYKQIAKHSFLASRHYI"
946..2370
/note="ORF2; ACMNPV ORF9 homolog"
/codon_start=1
/product="1629-capsid"
/protein_id="AAC59001.1"
/db_xref="GI:2934904"
/translation="MWERQYQSVRSYLNNHNAATAGPFLQRVAGPEAHSVGNVGD
RAVRLNQAVLDDLKLAEDIADYAYMOADQPEASSRHFATLNMRLLLIGVQDPDAR
RLNLSVLAIEALLRDVVDNDAENVLSGDFEYESKYISYQOFPAQTPTASQOOTO
TSLRPQTSLPQTOASLPQTPDQPEMVPSPFVHTTTPALLPQTPQPTADTFSRP
SDEYVYVGGKRAVPDTRFKPPPKPEHLKSRPSVATNAAGATVPVAPPPPPFASD
VTSMPPPPPPFASDVTSMPPPPPPMVDLATSMPPPPPPMVDLATSMPPPPINNA
INNLIDAMVAETNKNAGDNRSAALLDQIKQKTLAKTPADGAPATDPRSTLLSIRQ
CKTLKLRKIDQSSSTOTLLKDVDTDKTKTKTILKNFTVNIIDRISKQEQEKORLDTIT
KRRPAVEHTGDSNGNSDMDRDD"
complement(2533..3270)
/note="ORF3; ph, ACMNPV ORF8 homolog"
/codon_start=1
/product="polyhedrin"
/protein_id="AAC59002.1"
/db_xref="GI:1911249"

/translation="MPDYSYRPTIGRTYVYDKNYKKNLGSVILKNAKRRKHLLEBEDE
KHLDPDHYMVAEDPFLGPKGNKOKLTLEKEIRNVKPDYMKLILVNSCKEFLRTWTRF
VEDSPVINDQEVMDVFLVNNRPNRNCYKFLAQHALRWDCDIYPHEVIRLYEPSY
VGMNNEYRISLAKGGGCPIMNIAEYTNSTSFVNRVITWENFYKPIYIIGTDSSEE
EILIEVSLVFKVEFAPDAPLFTGPAY"
complement(3350..3541)
/note="ORF4"
/codon_start=1
/product="unknown"
/protein_id="AAC59003.1"
/db_xref="GI:1911250"
/translation="MRKNNGASQYPEHLTTVHKHRLQLSCVWCDDORSWDHPSAKGL
MFEHMFCCRLRIYVOOKIKH"
complement(3698..4093)
/note="ORF5"
/codon_start=1
/product="unknown"
/protein_id="AAC59004.1"
/db_xref="GI:1911251"
/translation="MGASONSEQILKLFYRWSOTGAALDDEKDLHCLYDLERFVGA
HLNKRADDKARKKCAEKAKALRVEIADRHLEAAAAACACADDGGRWSTLSRAQLDDI
AREKEIVDRILRLQKODSLLKQDLRKR"
complement(4125..4739)
/note="ORF6; lcf-2, ACMNPV ORF6 homolog"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAC59005.1"
/db_xref="GI:1911252"
/translation="MERVNPAGIDGLKRSETVLDPHDFVCLTSLPYTVFERGLF
VRMSGMRLLALLAAPQEPQPAVRFRFORSRRNVCLKACADGASLAAVAVSWP
PCMSKTMADLSAPRGNNYKRFNCLANVTCTCKTACGLIGALLHFTMDAKCV
GEVTHULLKAQDVKPSNCAKMKKVTKLCPQASMKCKGLNPGNF"
complement(4742..4975)
/note="ORF7; ACMNPV ORF5 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59006.1"
/db_xref="GI:1911253"
/translation="MNRTRMTAANTDYDREQLRRLNLSRRSVHCLTTSATGPD
CNFLAGDRAPAVIVKAAANGHQSLLCDKV"
5020..5460
/note="ORF8; ACMNPV ORF4 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59007.1"
/db_xref="GI:1911254"
/translation="MPGARFVRESLRLTOEFKENVVAHVHMLGLRALIDGKVTSDV
RRFGLSRNVLSACMNVVQYAPDATTDMRHOPTIYFRVCONCHAMADVAPPDHS
IARYLAECAVLVIDHPLDVFGTEGVNELLELVEQRLNAGGDL"
complement(5669..6151)
/note="ORF9; ptp-2"
/codon_start=1
/product="protein tyrosine phosphatase 2"
/protein_id="AAC59008.1"
/db_xref="GI:1911255"
/translation="MYEASQIDEHVFVGGYGDNEAMLRFTENHATASVILSDSDVA
PIRALGLPVGDRHIHYCEAAPTCAALPNAMPALYDMVRIGEGKRVLVHCVAGASR
SAAAYILNRSQMPYQDALNAVQSRRAVINDHIEVFLATRCYSRYFNDELKQIV"
complement(6129..6791)
/note="ORF10; ptp-1, ACMNPV ORF1 homolog"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAC59009.1"
/db_xref="GI:1911256"
/translation="MLTSRRVAYITTSERNCGCAWCTNGVRBRROVHLPSVARAM
FPDRMHYTCAGAVIEGTRLLCFVKVPLNAELFEYVTSDEDRWTAASVLAHSALCAVI
DLTNTARYYDGAQVMKGLLYKIRVPGRAVDDDIIVAEFIEYVDEFFRCPTMLVAV
HWTHLNRSGYLCVRYMVERLGVSPDTAIAIRFETARGHKIETNYLQDLARKVRQG
PN"
6819..7814
/note="ORF11; ACMNPV ORF11 homolog"
/codon_start=1

```

```
/product="unknown"
/protein_id="AAC59010.1"
/db_xref="GI:1911257"
/translation="MSLSKLLVAYAYGYSNLPDRYGESYLYRIVIEHLTNTYVSN
ASCVRDITACRLNSCHLCEVAVOLLDSVEAARLSAMFRCGDATGLCADMORALA
DIDRHAPLARVRGRANIEFALDAIADIPSDVTNNLOGIIGRPMHFPKCSGLARVADV
DPDIRADGWYHFKCVLTVMHLVACGAVPAGSATRLRDARAKHIGPNDEGNCAPATAA
VYGRCAIGREHFAHKTACMHILFQPMRNDLTPADERPCFVIGKDFGRQCKDTYTD
LRTHADALYIHGTTDRQKNALEFLDCCVNASDIDADCYDCVNVNFYATQNKYKM"
CDS
complement(8180..9142)
/notes="ORF12; AcMNPV ORF13 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59011.1"
/db_xref="GI:1911258"
/translation="MESNMGWNADEQVNAEFDEQAYRYRYVQDRAHSDLVRRDVF
RCPFVFRFRLVDETAGRCSSVDFCKGLKISHDLQRCNFRDQHQHROLNELVFGP
PAQPDGSLFATKGLVQLQOFSEFANKNEVLLAVGANKDHRDNLDDKILPVLNVH
KTLNDSKFIISAKHFKLEVGAQFEQRIQTLDTKLNALOCAAPTPTAPCVVPR
DYTHPHLAVFMGRVEDRCVQIAFARGQEHKRLKLEFEGMDVDVGRAPNPLLA
VHICKEEFANGHKKIRLPKPKVIEVCANVAKDIVKKAILNKI"
CDS
complement(9094..9825)
/notes="ORF13; 1ef-1, AcMNPV ORF14 homolog"
/codon_start=1
/product="late expression factor 1"
/protein_id="AAC59012.1"
/db_xref="GI:1911259"
/translation="MAPCKYTPERVOMMDATAYNDSRRLAFMTDRPRVHAHNEFDS
AAQLFAYIVKNSISDVHKVPLEEGREWVIDADFKDCAKAEMLKLVNKGATAPMLFP
EGKEDAVORIMFSGNRGPHLWKEGFKMDAPKSLREHFNVPKOPAKLVSGDIRPG
SFADCVRRVAVMYIGDAREDLVRYMPDVRDVFNCNANKQIRAPFSYNTAGGDISRCL
TOOLQORIKACAGSLAGGAGTPTSK"
CDS
9893..11362
/notes="ORF14, egt, AcMNPV ORF15 homolog"
/codon_start=1
/product="ecdysteroid UDP-glucosyltransferase"
/protein_id="AAC59013.1"
/db_xref="GI:1911260"
/translation="MVFLIALTLATGARAASILAVLPTPAYSHVYVYVAVHALYK
NCHNVTIKPOLLDYAVODEGVRVEIDADMSAOYKKLIVASSGVFRKGVVADETIV
TADNYGLTMEFKDQFNANVRELSTNETADVVEAFADYALVFGHLFRPFIQI
APGVGLAEFERERRAVAPHLVPTGAAALIRRGALSEWRLNEFELLARRSDELL
KQPGKSTPTIRQLDNDVQLLLNHPVYDNNRPPVSVQVYLGGLHLAQALPQLDA
PLERLINESVGAVIYVSGSDITNSIHAFLQMLDITFANLNNYTVLMKWDDAAS
VALPRNLQAKWFSQTVLRHKNVAVFTVQAGLSSDEALQARVPMVCLPMMGDOFHH
ARKLQDFGARALDTAAVSAPQLQAIREVIAADGEAYRARIKLRVAVHEHDAAPDEKA
VKTERVIRKNDNVNPFARSLKTTAANMAYSDYFVRPL"
CDS
11547..12140
/notes="ORF15; AcMNPV ORF16 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59014.1"
/db_xref="GI:1911261"
/translation="METAQPPISYAPPKRGAVAYRVTVTVTVSDSGNEDRLTO
IVAOQLRLNFKLSQLORRRVNMOKLIRKNSVIANLARLTOKTKKHFAVTR
KNYHTTSGSQFVRVLELCANGGEQVFCARRADCDRRRVAEALATALAGVVA
SAANKRFEIIDEKLVSAKLVQVQLHDGSDTCD"
CDS
12109..12732
/notes="ORF16; AcMNPV ORF17 homolog"
/codon_start=1
```

```
Query Match 3.1%; Score 46; DB 80; Length 131993;
Best Local Similarity 44.3%; Pred. No. 1.9;
Matches 187; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

Qy 756 aactatggtatcttagcaagtccattcttcgagccgcaacgagagagcggtgctcgg 815
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75798 AACAAAGCCACCTTTCCTCGCGACGCGACGCCAACGCCATCGCGCGCGACGCC 75739

Qy 816 ctcaactacagcaccaggaattcgaaccggtattctcttcttccactacgttc 875
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75738 AACGCCATCGCGACGCCGCCAACGCCGCGACGCCAACGCCGCGACGCC 75679
```

```
Qy 876 aaccagacctcccggtccaaacggtggtcccggttgaaacctttcttcaagaagcgctaccc 935
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75678 ATCCGCCGACGCCGCGACGCGAGCGGCTATCACCAACGCCGACGCCATCGCGCGACGCTATACC 75619

Qy 936 ggtgtctctccctaaattctcttaccattaaacctctctgctgagcaccacaacaaactcc 995
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75618 AACGCCATCACCAACGCCGCGACGCCATCGCGCGAGCGCCATCGCGCGACGCCATCACCAACGCC 75559

Qy 996 atgacacccacacgggtaacggtctctactacggtgaggtcttctcgtaagcagatctatgatgcc 1055
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75558 GACGCCATCGCGCGACGCCATCACCAACGCCGCGACGCCATCGCGCGACGCCATCACCAACGCC 75499

Qy 1056 attcgccaggggccctcagtgaggaacagacctgtcttcttaccattaccacgagagaccggt 1115
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75498 GACGCCATCACCAACGCCGCGACGCCATCGCGCGACGCCATCACCAACGCCGCGACGCCATCACCAACGCC 75439

Qy 1116 ggtctttagacacatgtctctctccctctctctctctctctctctctctctctctctctact 1175
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75438 GAGGCCATCGCGCGACGCCATCACCAACGCCGCGACGCCATCACCAACGCCGCGACGCCATCACCAACGCC 75379

Qy 1176 ga 1177
||
Db 75378 GA 75377
```

Search completed: February 27, 2001, 01:03:18  
Job time: 15821 sec



2

Toxoplasma oocyst; ss.

2  
3











PF 15-JUL-1996; 96WO-EP03091.  
XX 22-DEC-1995; 95EP-0870137.  
PR 14-JUL-1995; 95EP-0870089.  
XX (INNO-) INNOGENETICS NV.  
PA Scarcez T, Van Broekhoven A;  
PI WPI; 1997-132645/12.  
XX New nucleic acid encoding Saccharomyces pectinase - useful in food  
PT processing, also its regulatory sequences are useful for expression  
PT of heterologous protein  
XX Example 1; Fig 2; 65pp; English.  
XX This sequence represents the insert from cosmid 109, containing  
CC Saccharomyces cerevisiae strain FY1679 DNA. This sequence contained the  
CC pectinase gene of the invention (see T80042). Pectinases are synthesised  
CC by plants and a variety of microorganisms, the best characterised  
CC pectinase being polygalacturonase. Pectinases play roles in the invasion  
CC of plant tissues by phytopathogens, the spoilage of fruits and  
CC vegetables, and also in their food processing and plant biological  
CC applications. Vectors containing the pectinase coding sequence are used  
CC for production of homologous or heterologous proteins in transformed host  
CC cells. The pectinase sequence may include regulatory and/or coding  
CC sequences, including the sequence encoding the pectinase signal peptide  
CC (which can be used in any recombinant system). The pectinase is useful in  
CC food processing (e.g. preparation of fruit juices), typical heterologous  
CC proteins that can be made are (poly)peptides of bacterial, animal or  
CC viral origin. Also, probes derived from the pectinase coding sequence can  
CC be used to detect the coding sequence and primers for amplification are  
CC used to isolate related genes.  
XX Sequence 40875 BP; 12953 A; 7579 C; 8292 G; 12051 T; 0 other;

Query Match 1.9%; Score 44.6; DB 18; Length 40875;  
Best Local Similarity 43.9%; Pred. No. 0.028;  
Matches 191; Conservative 0; Mismatches 244; Indels 0; Gaps 0;  
Qy 1306 atgacgagcttctggaactatggtatctctagaagtcctatttcgagcgccaagc 1365  
Db 24625 ATCCCAACCTCTACTTACACGAGCTACAAATCTAGTACTTCTGACTTCTTACCACC 24566  
Qy 1366 agaaggcggtgctgctcaactacgatggcaccacaggaagattcgacccgattctc 1425  
Db 24565 ACTATCCTTCTACTTCTTACCACATCGACTACTTCTTACAACTCTTACTTCTTACC 24506  
Qy 1426 tcttttaactagtcacacgaactcccggtcccaacggtggccggttgaaactctt 1485  
Db 24505 ACTCTTACAACTTCTTACCACTCGACTACTTCTTACAACTTCTTACTTCTTACCCTCT 24446  
Qy 1486 tccaagacgcttactctggtgcttctcttaattcttcttaccattacccctctgctg 1545  
Db 24445 ACAACTTCTTACCACTCGACTTCTTCTTACAACTTCTTACTTCTTACCACTTCTTAC 24386  
Qy 1546 gcacaaac 1605  
Db 24385 TCTTACCACTCGACTTCTTCTTACAACTTCTTACTTCTTACCACTTCTTACTTCTTACA 24326  
Qy 1606 agcagatctatgatgcttcgcatcgcagggccctcagtgaggagacccctgctcttatta 1665  
Db 24325 ACTTCTACCACTTCTTACAACTTCTTACTTCTTACAACTTCTTACTTCTTACCACTTCT 24266  
Qy 1666 cctacagacgagcgggtgcttcttctacgacctgtctcctccctctcgcgctccgccc 1725  
Db 24265 CAACCTTCCACCAATCGACTTCTTCTTCTTCAAGCATCGACTTCTTCTTCACTTCTTCA 24206  
Qy 1726 acaacttgacctaca 1740  
Db 24205 TCTTACCACTTCTTACA 24191

RESULT 10  
Z61243  
ID Z61243 standard; DNA; 2110 BP.  
XX Z61243;  
AC 30-MAY-2000 (first entry)  
DT DNA encoding a phenol oxidising enzyme.  
DE Phenol oxidizing enzyme; fungus; redox reaction; detergent; bleaching;  
KW fabric; pulp; paper; decolourisation; plant-derived food product;  
KW coloured compound; porphyrin; tannin; polyphenol; carotenoid;  
KW anthocyanin; Maillard reaction product; ss.  
XX Acremonium murorum.  
OS Key Location/Qualifiers  
FH 135..1943  
FT /\*tag= a  
FT /product= "phenol oxidising enzyme"  
XX WO200005349-A1.  
XX 03-FEB-2000.  
XX 13-JUL-1999; 99WO-EP04922.  
XX 21-JUL-1998; 98EP-0202454.  
XX (UNIL ) UNILEVER NV.  
XX (UNIL ) UNILEVER PLC.  
XX (HIND-) HINDUSTAN LEVER LTD.  
XX Convents D, Gouka RJ, Van Der Heiden M, Swarthoff T, Verrips CT;  
XX WPI; 2000-195101/17.  
XX P-PSDB; Y69204.  
XX Phenol-oxidizing enzyme from Acremonium, used in detergent compositions  
XX for bleaching stains on fabrics -  
XX Example 2; Page 37-40; 45pp; English.  
XX The present sequence encodes a phenol oxidizing enzyme from the  
XX fungus Acremonium murorum. The enzyme has the CBS accession number  
XX 157.72. The enzyme catalyses redox reactions and is specific for  
XX molecular oxygen as the electron acceptor. The phenol oxidizing enzyme  
XX is specifically used in detergents for bleaching strains on fabrics,  
XX but also for bleaching pulp and paper and for decolourisation of  
XX plant-derived food products. The enzyme has a pH optimum in the  
XX alkaline to neutral range and can bleach a wide variety of coloured  
XX compounds, e.g. porphyrins, tannins, polyphenols, carotenoids,  
XX anthocyanins and Maillard reaction products.  
XX Sequence 2110 BP; 400 A; 732 C; 556 G; 422 T; 0 other;

Query Match 1.9%; Score 44.2; DB 21; Length 2110;  
Best Local Similarity 45.6%; Pred. No. 0.0066;  
Matches 142; Conservative 0; Mismatches 163; Indels 0; Gaps 0;  
Qy 1347 atcttgaagcgccgaacgagaggcggtgcttctgaactacgagcaccacgga 1406  
Db 1182 accatcgagctccgcaacgagccggtcggtcggtcggtcggtcggtcggtcggtcggtcg 1241  
Qy 1407 gaattcgacggaggtcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1466  
Db 1242 gacacgacaaaggtcatgcttcttcttcttcttcttcttcttcttcttcttcttctt 1301  
Qy 1467 gtgcccgttgaaacttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1526

```

Db 1302 tccgtctccctccaccctccgtgacgtgcccctccctccagcaccctcgaccaccatc 1361
Qy 1527 attaacccctctctgctgcggcaaccaacacacacacacacacacacacacacacacac 1586
Db 1362 gaccactgcttcgctgcccgcacccgcgcagtcagtcagtcagtcagtcagtcagtcac 1421
Qy 1587 tacagtgaggtcttcgcaagcagatctatgcccattccagtcagtcagtcagtcagtcag 1646
Db 1422 tccgacgtcgagaaacgcgcgtgctgccaacgctgcccctggaacgcgcagtcagtcag 1481
Qy 1647 aagac 1651
Db 1482 ctgac 1486

RESULT 11
T37329
ID T37329 standard; DNA; 28804 BP.
XX AC T37329;
XX
XX 30-NOV-1996 (first entry)
XX DE Sphingon biosynthetic gene region.
XX KW Sphingon; polysaccharide; spsB gene; glucosyl-IP-transferase; ds.
XX OS Sphingomonas strain S88 (ATCC 31554).
XX FH Key Location/Qualifiers
FT CDS complement (1942..1944)
FT FT /*tag= a
FT /*codon_start= 1942..1944
FT /*note= "spsG gene putative initiation codon"
FT /*tag= b
FT /*codon_start= 3311..3313
FT /*note= "spsS gene putative initiation codon"
FT /*tag= c
FT /*codon_start= 5323..5325
FT /*note= "spsR gene putative initiation codon"
FT 5526..5528
FT /*tag= d
FT /*codon_start= 5526..5528
FT /*note= "spsQ gene putative initiation codon"
FT /*tag= e
FT /*codon_start= 7076..7078
FT /*note= "spsI gene putative initiation codon"
FT 7588..7590
FT /*tag= f
FT /*codon_start= 7588..7590
FT /*note= "spsK gene putative initiation codon"
FT 8643..8645
FT /*tag= g
FT /*codon_start= 8643..8645
FT /*note= "spsL gene putative initiation codon"
FT /*tag= h
FT /*codon_start= 10938..10940
FT /*note= "spsJ gene putative initiation codon"
FT 11569..11571
FT /*tag= i
FT /*codon_start= 11569..11571
FT /*note= "spsF gene putative initiation codon"
FT 12886..12888
FT /*tag= j
FT /*codon_start= 12886..12888
FT /*note= "spsD gene putative initiation codon"
FT 15165..15167
FT /*tag= k

```

```

FT FT /codon_start= 15165..15167
FT FT /*note= "spsE gene putative initiation codon"
FT FT 15883..15885
FT FT /*tag= l
FT FT /*codon_start= 15883..15885
FT FT /*note= "urf32 gene putative initiation codon"
FT FT 16748..16750
FT FT /*tag= m
FT FT /*codon_start= 16748..16750
FT FT /*note= "urf26 gene putative initiation codon"
FT FT complement (21082..21084)
FT FT /*tag= n
FT FT /*codon_start= 21082..21084
FT FT /*note= "atrB gene putative initiation codon"
FT FT 21706..23118
FT FT /*tag= o
FT FT /label= spsB
FT FT /product= glucosyl IP-transferase
FT FT 23238..23240
FT FT /*tag= p
FT FT /*codon_start= 23238..23240
FT FT /*note= "rthA gene putative initiation codon"
FT FT 24113..24115
FT FT /*tag= q
FT FT /*codon_start= 24113..24115
FT FT /*note= "rthC gene putative initiation codon"
FT FT 24683..24685
FT FT /*tag= r
FT FT /*codon_start= 24683..24685
FT FT /*note= "rthB gene putative initiation codon"
FT FT 25744..25746
FT FT /*tag= s
FT FT /*codon_start= 25744..25746
FT FT /*note= "rthD gene putative initiation codon"
FT FT complement (27534..27536)
FT FT /*tag= t
FT FT /*codon_start= 27534..27536
FT FT /*note= "urf31 gene putative initiation codon"
FT FT 27747
FT FT /*tag= u
FT FT /*codon_start= 27747..27749
FT FT /*note= "urf34 gene putative initiation codon"
FT XX EP728841-A2.
FT XX
FT XX 28-AUG-1996.
FT XX
FT XX 24-JAN-1996; 96EP-0300467.
FT XX
FT XX 24-JAN-1995; 95US-0377440.
FT XX (SHIN-) SHINETSU BIO INC.
FT XX (SHIE ) SHINETSU CHEM CO LTD.
FT XX Armentrout RW, Mikolajczak M, Pollock TJ, Thorne L;
FT XX Yamazaki M;
FT XX WPI; 1996-386292/39.
FT XX P-PSDB; W03997.
FT XX
FT XX New isolated DNA from Sphingomonas sp. - used for transforming
FT XX recipient bacteria to obtain hyper-producers of sphingon
FT XX polysaccharide(s).
FT XX
FT XX Claim 32; Page 56-70; 105pp: English.
FT XX
FT XX A 28.8 kb chromosomal fragment of Sphingomonas strain S88 was
FT XX isolated on the basis of its ability to restore sphingon
FT XX biosynthetic capability to Sphingomonas mutant S88m260. It
FT XX contains 23-25 genes, including sps genes coding for biosynthesis of
FT XX the polysaccharide sphingon, rth genes coding for drdp-(L)-hamnose
FT XX biosynthesis, atrDB genes coding for a transport function and some
FT XX unidentified open translation reading frames (urf). The spsB gene

```





CC The sequence is that of an example randomising oligonucleotide which  
 CC is used in the prepn. of mRNA encoding candidate polypeptides for the  
 CC method of systematic polypeptide evolution by reverse translation  
 CC (SPERT). The method provides a rapid way of isolating and identifying  
 CC polypeptide ligands which bind to target mols. The polypeptide ligands  
 CC can be used in e.g. assay methods, diagnostic procedures, cell sorting,  
 CC as activators or inhibitors of target mol. function, as probes, as  
 CC sequestering agents, drug delivery vehicles, modifiers of hormone  
 CC action and as catalysts. See also Q21830-Q21832.

XX  
 SQ Sequence 390 BP; 125 A; 126 C; 133 G; 6 T; 0 other;

Query Match	1.8%	Score 41.2;	DB 13;	Length 390;
Best Local Similarity	44.7%	Pred. NO. 0.02;		
Matches 160;	Conservative 0;	Mismatches 198;	Indels 0;	Gaps 0;

Qy	1343	gtccatcttcgagccgccaacgagagggcggtgctggtcgaactacgacgatggaccaa	1402
Db	2	ggccatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacga	61
Qy	1403	cggagaattcgaccggattctcttcttcacctacgtcaaccagacctcccggtccaa	1462
Db	62	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacga	121
Qy	1463	cgtggtgcccgttgaaacttctccaagacgacctacctcgtgctccctccataattctc	1522
Db	122	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacga	181
Qy	1523	ttacattaaacctctctgctgctgacacacacaaactccatgaccccccgcgtaacgt	1582
Db	182	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacga	241
Qy	1583	ctctacggtgaggtcttcgcaagcagatctatgatgcattccgagggccctcagtg	1642
Db	242	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacga	301
Qy	1643	ggacaagacctgctcttcattacctacgacgacgacggtggcttctacgacctatgc	1700
Db	302	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacga	359

Search completed: February 26, 2001, 14:08:46  
 Job time: 42672 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2001, 23:26:55 ; Search time 2132.85 Seconds  
(without alignments)  
7615.799 Million cell updates/sec

Title: US-09-426-072-1

Perfect score: 2318

Sequence: 1 aattcttgagacaataatccc.....gcacccggatcacatttaatt 2318

Scoring table: IDENTITY\_NUC

,Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estov1.\*  
68: em\_estov2.\*  
69: em\_estpl1.\*  
70: em\_estpl2.\*  
71: em\_estpl3.\*  
72: em\_estpl4.\*  
73: em\_estpl5.\*  
74: em\_estrol.\*  
75: em\_estro2.\*  
76: em\_estro3.\*  
77: em\_estro4.\*  
78: em\_estro5.\*  
79: em\_estro6.\*  
80: em\_estro7.\*  
81: em\_estro8.\*  
82: em\_estro9.\*  
83: em\_estro10.\*  
84: em\_estro11.\*  
85: em\_estro12.\*  
86: em\_estro13.\*  
87: gb\_est41.\*  
88: gb\_est42.\*  
89: gb\_est43.\*  
90: gb\_est44.\*  
91: gb\_est45.\*  
92: gb\_est46.\*  
93: gb\_est47.\*  
94: gb\_est48.\*  
95: gb\_est49.\*  
96: gb\_est50.\*  
97: gb\_est51.\*  
98: gb\_est52.\*  
99: gb\_est53.\*  
100: gb\_est54.\*  
101: gb\_est55.\*  
102: gb\_est56.\*  
103: gb\_est57.\*  
104: gb\_est67.\*  
105: gb\_est68.\*  
106: gb\_est69.\*  
107: gb\_est70.\*  
108: gb\_est71.\*  
109: gb\_est72.\*  
110: gb\_est73.\*  
111: gb\_est74.\*  
112: em\_esthum21.\*  
113: em\_esthum22.\*  
114: em\_esthum23.\*  
115: em\_estom1.\*  
116: em\_estom2.\*

117: em\_estpl6:\*  
 118: em\_estpl7:\*  
 119: em\_estpl8:\*  
 120: em\_estpl9:\*  
 121: em\_estpl10:\*  
 122: em\_estpl11:\*  
 123: em\_estpl12:\*  
 124: em\_estpl13:\*  
 125: em\_estpl14:\*  
 126: em\_estpl15:\*  
 127: em\_estpl16:\*  
 128: em\_estpl17:\*  
 129: em\_estpl18:\*  
 130: em\_estpl19:\*  
 131: em\_estpl20:\*  
 132: em\_estpl21:\*  
 133: em\_estpl22:\*  
 134: em\_estpl23:\*  
 135: em\_estpl24:\*  
 136: em\_estpl25:\*  
 137: em\_estpl26:\*  
 138: em\_estpl27:\*  
 139: em\_estpl28:\*  
 140: em\_estpl29:\*  
 141: em\_estpl30:\*  
 142: em\_estpl31:\*  
 143: em\_estpl32:\*  
 144: em\_estpl33:\*  
 145: em\_estpl34:\*  
 146: em\_estpl35:\*  
 147: em\_estpl36:\*  
 148: em\_estpl37:\*  
 149: em\_estpl38:\*  
 150: em\_estpl39:\*  
 151: em\_estpl40:\*  
 152: em\_estpl41:\*  
 153: em\_estpl42:\*  
 154: em\_estpl43:\*  
 155: em\_estpl44:\*  
 156: em\_estpl45:\*  
 157: em\_estpl46:\*  
 158: em\_estpl47:\*  
 159: em\_estpl48:\*  
 160: em\_estpl49:\*  
 161: em\_estpl50:\*  
 162: em\_estpl51:\*  
 163: em\_estpl52:\*  
 164: em\_estpl53:\*  
 165: em\_estpl54:\*  
 166: em\_estpl55:\*  
 167: em\_estpl56:\*  
 168: em\_estpl57:\*  
 169: em\_estpl58:\*  
 170: em\_estpl59:\*  
 171: em\_estpl60:\*  
 172: em\_estpl61:\*  
 173: em\_estpl62:\*  
 174: em\_estpl63:\*  
 175: em\_estpl64:\*  
 176: em\_estpl65:\*  
 177: em\_estpl66:\*  
 178: em\_estpl67:\*  
 179: em\_estpl68:\*  
 180: em\_estpl69:\*  
 181: em\_estpl70:\*  
 182: em\_estpl71:\*  
 183: em\_estpl72:\*  
 184: em\_estpl73:\*  
 185: em\_estpl74:\*  
 186: em\_estpl75:\*  
 187: em\_estpl76:\*  
 188: em\_estpl77:\*  
 189: em\_estpl78:\*

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	119.6	5.2	667	92	AW565749	AW565749 LG1_349_G
2	74.4	3.2	466	107	BE415670	BE415670 MWE037.HO
3	74.4	3.2	472	110	BE586656	BE586656 WHE0509.A
4	73.6	3.2	617	107	BE404795	BE404795 WHE1205_E
5	70.2	3.0	561	135	BE758821	BE758821 en_3066_A
6	63.6	2.7	480	83	AW266416	AW266416 L30-3180T
7	62.6	2.7	601	88	AW266012	AW266012 L30-2674T
8	62	2.7	800	108	BE430717	BE430717 SUN007.E0
9	61	2.6	698	94	AW726682	AW726682 GA_Ea002
10	61	2.6	865	134	BE053545	BE053545 GA_Ea003
11	59.8	2.6	661	9	AA556789	AA556789 631 Loblo
12	58.6	2.5	392	138	BF010466	BF010466 NXCI_084
13	57.2	2.5	631	23	AL691435	AL691435 606015H11
14	56.8	2.5	490	94	AW756470	AW756470 SL21905.Y
15	53.6	2.3	493	134	BE057023	BE057023 945037H04
16	51.8	2.2	581	24	AL731956	AL731956 BNUGH115
17	51.6	2.2	380	145	T44962	T44962 8725 Lambda
18	51	2.2	487	27	AL994923	AL994923 701500427
19	50.2	2.2	676	19	AL389106	AL389106 GH20192.5
20	48.8	2.1	514	138	BF004644	BF004644 EST433142
21	48.8	2.1	552	138	BF004645	BF004645 EST433143
22	46	2.0	452	87	AW224301	AW224301 EST301028
23	46	2.0	545	108	BE436321	BE436321 EST407399
24	46	2.0	596	97	AW929343	AW929343 EST338131
25	45.8	2.0	400	139	C26175	C26175 C26175 Rice
26	45.6	2.0	500	174	B67199	B67199 CGS0015B Cp
27	45.2	1.9	470	29	AU095576	AU095576 AU095576
28	44.2	1.9	461	29	AU032411	AU032411 AU032411
29	44.2	1.9	829	107	BE413313	BE413313 MCG012.F0
30	43.8	1.9	466	134	BE051051	BE051051 za71907.b
31	42.8	1.8	373	24	AL737037	AL737037 sb36h07.Y
32	42.8	1.8	774	134	BE040321	BE040321 OE03H03.O
33	42.4	1.8	742	38	AW674403	AW674403 AV674403
34	42.2	1.8	553	87	AW180278	AW180278 MGA0370F
35	42.2	1.8	574	26	AL861086	AL861086 603011G12
36	42	1.8	445	133	BE012244	BE012244 945015H09
37	41.8	1.8	365	169	A2242255	A2242255 RPCI-23-9
38	41.6	1.8	457	109	BE516820	BE516820 WHEG20.D0
39	41.6	1.8	537	27	AL948319	AL948319 603042C12
40	41.6	1.8	1101	190	CNS00F7H	AL070235 Drosophil
41	41.4	1.8	891	192	CNS0416L	AL269814 Tetraodon
42	40.8	1.8	515	109	BE519721	BE519721 HV_CEB001
43	40.6	1.8	681	191	CNS02EOD	AL193900 Tetraodon
44	40.4	1.7	436	29	AU068299	AU068299 AU068299
45	40.4	1.7	442	144	R59827	R59827 yhlle03.F1

ALIGNMENTS

RESULT 1  
 AW565749  
 LOCUS LG1\_349\_G03.g1\_A002 Light Grown 1 (LG1) Sorghum bicolor CDNA, mRNA  
 DEFINITION sequence.  
 AW565749  
 ACCESSION  
 VERSION  
 KEYWORDS EST.  
 SOURCE sorghum.





```

Db 241 AGCTGCCAATACGTCGTCA---TCGAGCAGCGCTACTTCGATTGTCGAAGGAGTTCGCCG 297
Qy 1555 ccaactcattgacccccaccggttaacgtctctcctacggtgaggtctctcgtcaagcaatct 1614
Db 298 CCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 357
Qy 1615 atgatgctattccagggccctcagtgaggaagaacacctctcttcttaccctacgacg 1674
Db 358 AGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 417
Qy 1675 agacccgttgcttctacacattgtccctccctccctccg 1715
Db 418 AGCATGGTGGCTTCTATGACCATGTCCCCACGCGCTCGTC 458

```

## RESULT 3

```

LOCUS BE586656 472 bp mRNA EST 17-AUG-2000
DEFINITION WHE0509_A09_A17R Secale cereale aluminum-stressed root tip cDNA
ACCESSION BE586656
VERSION BE586656.1 GI:9839688
KEYWORDS EST.
SOURCE rye.
ORGANISM Secale cereale
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Secale.
REFERENCE 1 (bases 1 to 472)
AUTHORS Anderson,O.D., Butler,E., Chao,S., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - Aluminum-stressed root tip cDNA library from rye (Secale cereale)
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: M13 reversed primer.
FEATURES
Location/Qualifiers
1..472
/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE0509_A09_A17"
/clone_lib="Secale cereale aluminum-stressed root tip cDNA library"
/tissue_type="Root tip"
/lab_host="E. coli DH12S"
/lab_stage="Seedling"
/notes="Vector: pSPORT1; Site_1: Sali; Site_2: NotI; Seeds were germinated aseptically on filter paper and transferred to a hydroponic growth system in a growth chamber when the primary root was 1 cm in length. After a 2-day establishment period, seedlings were subjected to a 5 ppm aluminum stress prior to tissue harvest. Plants were grown in an environmental chamber. The tissue, total RNA, and poly(A) RNA were prepared, and a cDNA library was made (Butler and Gustafson) at University of Missouri, Columbia. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

```

BASE COUNT  
ORIGIN

94 a 164 c 120 g 94 t

```

Query Match 3.2%; Score 74.4; DB 110; Length 472;
Best Local Similarity 65.9%; Pred. No. 9.2e-11;
Matches 108; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 1553 caccacatccatgacccccaccggttaacgtctctcctacggtgaggtctctcgtcaagcaatct 1612
Db 219 CGCCACGACGACCCCTCGCAGCAGCTCGCAGGGCCGACGAGGTTTGTCAAGGAGGT 278
Qy 1613 ctatgatgctattccagggccctcagtgaggaagaacacctctcttcttaccctacgacg 1672
Db 279 CTACGAGACGCTGCGAGCAGCGCAGTGAATGAGACTCTCTCATCATCATCATATGA 338
Qy 1673 cgagaccggtggtctctacgacattctctccctccctccg 1716
Db 339 TGAGCATGTGGCTTCTACGACCATGTCCCCACGCGCTCACCG 382

```

## RESULT 4

```

LOCUS BE404795 617 bp mRNA EST 21-JUL-2000
DEFINITION WHE1205_E03_I05ZS Wheat etiolated seedling root cDNA library
ACCESSION BE404795
VERSION BE404795.1 GI:9364263
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Triticum.
REFERENCE 1 (bases 1 to 617)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Strategene SK primer.
FEATURES
Location/Qualifiers
1..617
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1205_E03_I05"
/clone_lib="Wheat etiolated seedling root cDNA library"
/tissue_type="Root"
/lab_host="E. coli SOLR"
/lab_stage="Five day old etiolated seedling"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, mystatin and cefotaxime in covered crystallization dishes. Roots were harvested. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TX Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

```

BASE COUNT  
ORIGIN

141 a 202 c 151 g 123 t

Query Match

3.2%; Score 73.6; DB 107; Length 617;









Tue Feb 27 08:38:59 2001

COMMENT

Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 945037 row: H column: 04.

FEATURES

Location/Qualifiers  
1..493  
/organism="Zea mays"  
/cultivar="W23"  
/db\_xref="taxon:4577"  
/clone\_lib="945 - Mixed adult tissues from Walbot lab,  
same as 707 (SK)"  
/tissue\_type="tassel, kernal, silk, husk, root, leaf"  
/dev\_stage="fully-grown"  
/lab\_host="DH10B"  
/note="Organ: tassel, kernal, silk, husk, root, leaf;  
Vector: pGAD10; Site: EcoRI; cDNA library from fully  
differentiated maize tissues from an active Mutator plant.  
Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,  
root, leaf). Unidirectionally cloned. New library number  
given to library 707 for additional sequencing."

BASE COUNT 114 a 150 c 130 g 99 t  
ORIGIN

Query Match 2.3%; Score 53.6; DB 134; Length 493;  
Best Local Similarity 60.1%; Pred. No. 0.00013;  
Matches 89; Conservative 0; Mismatches 59; Indels 0; Gaps 0;  
Oy 1566 caccaccggttaacgtctctacgtgaggtcttcgtcaagcagatcttatgatgccatt 1625  
Db 320 CATCGTCGCATGTCGCCGAGGGGGCACAGGCTTCGTGAAGAGGTGTATGAGACACTA 379  
Oy 1626 cggcaggccctcagtgaggacaagaccctctcttaccacacgagaccggtggc 1685  
Db 380 AGGGCCAGTCGGCAGTGAACGAGACTGCCCTTCATCATATGATGAGCATGTGTGA 439  
Oy 1686 ttctacgaccatgtccctccctctctg 1713  
Db 440 TTCTATGACCATGTACTACGCCGCTCG 467

Search completed: February 26, 2001, 04:51:53  
Job time: 19498 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2001, 20:39:37 ; Search time 4495.24 Seconds  
(without alignments)  
2639.001 Million cell updates/sec

Title: us-09-426-072-1

Perfect score: 2318

Sequence: 1 aattcttgacacaaatccc.....gcacccggatacataatt 2318

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_ph.\*  
6: gb\_pl1.\*  
7: gb\_pl2.\*  
8: gb\_pri.\*  
9: gb\_pr2.\*  
10: gb\_pr3.\*  
11: gb\_ro.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: em\_fun.\*  
15: em\_hum1.\*  
16: em\_hum2.\*  
17: em\_in.\*  
18: em\_om.\*  
19: em\_or.\*  
20: em\_ov.\*  
21: em\_pat.\*  
22: em\_ph.\*  
23: em\_pl.\*  
24: em\_ro.\*  
25: em\_sts.\*  
26: em\_sy.\*  
27: em\_un.\*  
28: em\_vi.\*  
29: gb\_ba3.\*  
30: gb\_in1.\*  
31: gb\_in2.\*  
32: gb\_in3.\*  
33: gb\_pl3.\*  
34: gb\_pr4.\*  
35: em\_ba1.\*  
36: em\_ba2.\*  
37: em\_htg1.\*  
38: em\_htg2.\*  
39: em\_htg3.\*  
40: em\_htg4.\*  
41: em\_htg5.\*  
42: em\_htg6.\*  
43: em\_htg7.\*

44: em\_htg8.\*  
45: em\_htg9.\*  
46: em\_htg10.\*  
47: em\_hum3.\*  
48: em\_hum4.\*  
49: em\_hum5.\*  
50: em\_hum6.\*  
51: gb\_pr5.\*  
52: gb\_pr6.\*  
53: gb\_pr7.\*  
54: gb\_htg1.\*  
55: gb\_htg2.\*  
56: gb\_htg3.\*  
57: gb\_htg4.\*  
58: gb\_htg5.\*  
59: gb\_htg6.\*  
60: gb\_htg7.\*  
61: gb\_htg8.\*  
62: gb\_htg9.\*  
63: gb\_htg10.\*  
64: gb\_htg11.\*  
65: gb\_htg12.\*  
66: gb\_htg13.\*  
67: gb\_htg14.\*  
68: gb\_htg15.\*  
69: gb\_htg16.\*  
70: gb\_htg17.\*  
71: gb\_htg18.\*  
72: gb\_htg19.\*  
73: gb\_htg20.\*  
74: gb\_htg21.\*  
75: gb\_htg22.\*  
76: gb\_htg23.\*  
77: gb\_sts1.\*  
78: gb\_sts2.\*  
79: gb\_vil.\*  
80: gb\_vil2.\*  
81: gb\_pat1.\*  
82: gb\_pat2.\*  
83: em\_htg0.\*  
84: gb\_htg24.\*  
85: gb\_pr8.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	64.8	2.8	82289	7	ATAC009895	AC009895 Arabidops
2	64.8	2.8	89934	7	ATAC009327	AC009327 Arabidops
C 3	55	2.4	96679	7	ATT8p19	AL133315 Arabidops
4	52.6	2.3	7218	81	I66494	I66494 Sequence 14
C 5	52.4	2.3	28291	31	CEP56H9	274473 Caenorhabd
C 6	52	2.2	30960	31	CEL46C8AL	AC024785 Caenorhab
7	50.6	2.2	2282	1	AF107252	AF107252 Burkholde
C 8	50.2	2.2	2426	60	AC020448	AC020448 Drosophil
C 9	50.2	2.2	260910	30	AE003831	AE003831 Drosophil
10	49.4	2.1	13669	80	TTV1TP	X14855 Thermoprote
11	49.2	2.1	104533	7	F10K1	AC067971 Sequence
12	47.8	2.1	4034	80	TTV1VP	X14717 Thermoprote
C 13	46	2.0	749	80	NPORPT2	D13755 Multicapsid
C 14	46	2.0	30941	29	SC4C6	AL079355 Streptomy
C 15	46	2.0	131993	80	OPU75930	U75930 Orgyia pseu
C 16	45.8	2.0	5500	7	ATGLYRP	X58338 A.thaliana
C 17	45.8	2.0	92620	6	AB026636	AB026636 Arabidops
C 18	45.6	2.0	144241	60	AC018929	AC018929 Oryza sat
C 19	45.6	2.0	147371	65	AC027037	AC027037 Oryza sat
C 20	45.4	2.0	5499	31	AF068065	AF068065 Cryptospo
21	45.4	2.0	108409	79	AF083424	AF083424 Atgeline h

22 45.2 1.9 1425 81 AR068771 Sequence  
 23 45.2 1.9 1425 81 AR092645 Sequence  
 24 44.8 1.9 2626 2 PSEPLC  
 25 44.8 1.9 11622 1 AE004519  
 26 44.8 1.9 35284 29 SCE19A  
 27 44.6 1.9 2783 7 CCLPGN1  
 28 44.6 1.9 9958 33 SCYJRL151C  
 29 44.2 1.9 41906 29 SC5C7  
 30 44.2 1.9 2126 7 AMU271104  
 31 44 1.9 750 32 SGSCRRHP  
 32 44 1.9 1419 33 FSU13050  
 33 44 1.9 11970 1 AE001978  
 34 44 1.9 28840 2 CNSPAX04  
 35 43.8 1.9 2603 7 COPEPCAX  
 36 43.8 1.9 64998 58 AC014035  
 37 43.8 1.9 269518 30 AE003751  
 38 43.6 1.9 35284 29 SCE19A  
 39 43.4 1.9 662 2 MSGMTP40  
 40 43.4 1.9 3013 2 MSGMTP40A  
 41 43.4 1.9 4094 2 MTU49511  
 42 43.4 1.9 31225 2 MTCY98  
 43 43.4 1.9 40892 81 AS9672  
 44 43.2 1.9 37521 29 SC1A8A  
 45 43.2 1.9 39446 29 SCB87

## ALIGNMENTS

RESULT 1  
 LOCUS ATAC009895 82289 bp DNA PLN 09-OCT-1999  
 DEFINITION Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence,  
 complete sequence.  
 ACCESSION AC009895  
 VERSION 2 GI:6017096  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 82289)  
 Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.,  
 Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E.,  
 Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.  
 Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence  
 Unpublished  
 2 (bases 1 to 82289)  
 Lin,X. and Kaul,S.  
 Direct Submission  
 Submitted (04-SEP-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org  
 3 (bases 1 to 82289)  
 Lin,X.  
 Direct Submission  
 Submitted (09-OCT-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 On Oct 9, 1999 this sequence version replaced gi:5822970.  
 Address all correspondence to:  
 Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr.  
 Rockville, MD 20850, USA  
 e-mail: xlin@tigr.org  
 BAC clone T21P5 is from Arabidopsis chromosome III and is near the  
 molecular marker ml172.  
 The orientation of the sequence is from SP6 to T7 end of the BAC  
 clone.  
 Genes were identified by a combination of three methods: Gene  
 prediction programs including GRAIL (available by anonymous ftp  
 from arthur.epm.ornl.gov), Genefinder (Phil Green, University of  
 Washington), Genscan (Chris Burge,

http://www.cbs.dtu.dk/netgene/cbsnetgene.html), and NetPlantGene  
 (http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the  
 complete sequence against a peptide database and the Arabidopsis  
 EST database at TIGR (http://www.tigr.org/tdb/at/at.html).  
 Annotated genes are named to indicate the level of evidence for  
 their annotation. Genes with similarity to other proteins are named  
 after the database hits. Genes without significant peptide  
 similarity but with EST similarity are named as 'unknown' proteins.  
 Genes without protein or EST similarity, that are predicted by more  
 than two gene prediction programs over most of their length are  
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
 predicted by tRNAscan-SE (Sean Eddy,  
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are  
 identified by RepeatMasker (Arian Smit,  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of  
 genomic sequence that are not annotated as genes but have predicted  
 exons by GRAIL are annotated as misc features.

FEATURES  
 source

1..82289  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="III"  
 /map="ml172"  
 /clone="T21P5"  
 1..24217  
 /note="Overlap with BAC clone T12J13  
 (AC009327:65718..89934)."  
 complement(join(<8..215,451..587,675..857,959..1050,  
 1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,  
 2979..3100))  
 /gene="T21P5.1"  
 complement(<8..>3100)  
 /gene="T21P5.1"  
 /note="similar to hypothetical protein GB:CA838918  
 [Arabidopsis thaliana]"  
 complement(join(8..215,451..587,675..857,959..1050,  
 1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,  
 2979..3100))  
 /gene="T21P5.1"  
 /note="unknown protein"  
 /codon\_start=1  
 /protein\_id="AAF01600.1"  
 /db\_xref="GI:6017117"

## misc\_feature

mRNA

## gene

## CDS

## mRNA

## gene

## CDS

join(<3627..3661,3762..3843,3918..3995,4081..4140,  
 4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,  
 5361..5494,5576..5682,5785..5827,5903..56144)  
 /gene="T21P5.2"  
 <3627..>6144  
 /gene="T21P5.2"  
 /note="predicted by genscan"  
 join(3627..3661,3762..3843,3918..3995,4081..4140,  
 4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,  
 5361..5494,5576..5682,5785..5827,5903..6144)  
 /gene="T21P5.2"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAF01580.1"  
 /db\_xref="GI:6017097"

/translation="MDDNRSSSEIKRHEIEKDTIASRKLEDTNTKLIQDPEEMALYAK  
 VRSEIEIHSIQERIAACLMOLLNEKYLKRCADLRVAIDEKQNEYSALNEL  
 ARRGDLEENKLAHDLKVTEDERYIFMTSLGLLAEYGVWPRVANATAISSGKHLH























structural proteins of the Thermoproteus tenax virus TTv1  
 Mol. Gen. Genet. 217 (1), 105-110 (1989)  
 89364672  
 3 (bases 12541 to 13242)  
 Neumann, H. and Zillig, W.  
 The TTv1-encoded viral protein TPX: primary structure of the gene  
 and the protein  
 Nucleic Acids Res. 18 (1), 195 (1990)  
 90174928

Location/Qualifiers  
 1. .13669 /organism="Thermoproteus tenax virus 1"  
 /strain="Kral"  
 /db\_xref="taxon:10479"  
 292. .489  
 /note="URF (65 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32969.1"  
 /db\_xref="GI:62154"  
 /db\_xref="SWISS-PROT:P19276"  
 /translation="MVDFDIEVESIPETMEEVGEYIDIFIQLTETAEMERDVIVEE  
 QKDEAKQFEYEDFWDVYT"  
 486. .863  
 /note="URF (127 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32970.1"  
 /db\_xref="GI:62155"  
 /db\_xref="SWISS-PROT:P19277"  
 /translation="MTRYNIRNLRIIRGEIKVGVSSDRIPMIEIPGHNDPERMCV  
 MTAEMEYSHIPIQADNLDEYDICKILTITVYINVAIGFGKNLGLKSEMLLVTRD  
 KKRFEKVGLELGIKYAYNWTCKFQ"  
 838. .1011  
 /note="URF (57 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32971.1"  
 /db\_xref="GI:62156"  
 /db\_xref="SWISS-PROT:P19278"  
 /translation="MHIIGVFSRVTFLDFYNYLQSGLYETLRIDRKYEYTLKMK  
 YATRLYGYVHT"  
 989. .1180  
 /note="URF (63 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32972.1"  
 /db\_xref="GI:62157"  
 /db\_xref="SWISS-PROT:P19279"  
 /translation="MASTFTHDVMECYEALAVMRCEVAYDRVLTWKMEPSPLLR  
 LDNEILIPYMKVYVYKQL"  
 1289. .1534  
 /note="URF (81 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32973.1"  
 /db\_xref="GI:62158"  
 /db\_xref="SWISS-PROT:P19280"  
 /translation="MAELNKAIKLTLYKLAMNYPYTADIAEKLELGLIRQEKWLKAI  
 NDTKRFKDKLDKRNNSNKIEKSENTGENHNDQQR"  
 complement(1467. .1688)  
 /note="URF (73 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32975.1"  
 /db\_xref="GI:62160"  
 /db\_xref="SWISS-PROT:P19281"  
 /translation="MRHAPSVLSRLFTIIVDFYDFWFSMRIPCTIWTYSISCTYGIYLL  
 ILIYRIFSGIDYVHDSLLYSILFLYC"  
 1509. .1850  
 /note="coat protein TP1 (AA 1 -113)"  
 /codon\_start=1  
 /transl\_table=11

protein\_id="CAA32974.1"  
 /db\_xref="GI:62159"  
 /db\_xref="SWISS-PROT:P19270"  
 /translation="MILIKTKNREYTDENKEINSISAADTICPYRAYRSYHRKIPJK  
 IDCKEARKIAGCSWTHYVILRQLKEQCETEKIVDPQLKFRADAICDGDALLRLQTS  
 IGTGMRLLISGN"  
 complement(1771. .2037)  
 /note="URF (88 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32977.1"  
 /db\_xref="GI:62162"  
 /db\_xref="SWISS-PROT:P19283"  
 /translation="MLHHPILFSLSTASPEFIRLIRFLTYSVIFTRSSSSSICIVLMP  
 LLMYIRCLPASAPASAIYTFENCHLLSACOSRLKPVLISTMRPHH1"  
 complement(1889. .2065)  
 /note="URF (58 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32978.1"  
 /db\_xref="GI:62163"  
 /db\_xref="SWISS-PROT:P19284"  
 /translation="MITFSKFAVYTPYSSIPLLHSFSIFYSSYQISNLSNFSYLLF  
 VHLICRPYIAVNVY"  
 1992. .2162  
 /note="URF (56 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32976.1"  
 /db\_xref="GI:62161"  
 /db\_xref="SWISS-PROT:P19282"  
 /translation="MEKMRGIEEGYGTAYANLKLAVIINCIEGERSHRDGNINIIHC  
 SAKPYSLUCET"  
 2074. .2286  
 /note="URF (70 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32979.1"  
 /db\_xref="GI:62164"  
 /db\_xref="SWISS-PROT:P19285"  
 /translation="MKAKEVIELEIIIIYVVPVSLIVVSHNTVYVAITVTLNL  
 ARKTYTKKIEKMKKVERALEFFS"  
 complement(2138. .2389)  
 /note="URF (83 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32981.1"  
 /db\_xref="GI:62166"  
 /db\_xref="SWISS-PROT:P19287"  
 /translation="MLPLMAAIISSSIGIGTSSYSPRVTLISLIVYTLMRKTRAPFL  
 LSSFFQSFLYKSSLSLNSMLPULLLHPIYVYKHNL"  
 2283. .2582  
 /note="URF (99 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32980.1"  
 /db\_xref="GI:62165"  
 /db\_xref="SWISS-PROT:P19286"  
 /translation="MRVTIRDDRIVYLGVEYELVPMKPLEDDIAAKKSITAYVLNSH  
 YSFIKALDKTMELENLAYELHONIDRTPTLTRIYLLMKVYGMHGSILSQF"  
 2672. .3076  
 /note="URF (134 AA)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA32982.1"  
 /db\_xref="GI:62167"  
 /db\_xref="SWISS-PROT:P19288"  
 /translation="MKFINLLMLMSVVVGRIAGANELEMYQAGIEVKRDVQVVR  
 VKKISYRAPRIADALNIYVEDKISGKPYIPGLTGLVKAIKIAYRDGVDSAMNI  
 FKSVDYAIQNEIRKIIIEGLMKVKEIRSEGRP"  
 3073. .3498  
 /note="URF (141 AA)"  
 /codon\_start=1





```

Qy 1806 accctgggttatccctccctacagtaagaagg 1837
||||| | | | | | | | | | | | | | | | | |
Db 17713 ACCCTTCTGATCTCCTCTGGATTGAGAAGG 17744

RESULT 12
LOCUS TTVPVP 4034 bp DNA 10-FEB-1999
DEFINITION Thermoproteus tenax virus 1 viral protein TPX-VT3 exhibiting
genomic variation.
ACCESSION X14717
VERSION X14717.1 GI:62191
KEYWORDS unidentified reading frame.
SOURCE Thermoproteus tenax virus 1.
ORGANISM Thermoproteus tenax virus 1.
VIRUSES: dsDNA viruses, no RNA stage; Lipothrixviridae;
Lipothrixvirus.
REFERENCE 1 (bases 1 to 4034)
AUTHORS Neumann, H.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-1989) Neumann H., Max-Planck-Institut fuer
Biochemie, Am Klopperspitz, 8033 Martinsried, F R G
REFERENCE 2 (bases 1 to 4034)
AUTHORS Neumann, H. and Zillig, W.
TITLE Nucleotide sequence of the viral protein TPX of the TTV1 variant
VT3
JOURNAL Nucleic Acids Res. 18 (8), 2171 (1990)
MEDLINE 90245666
COMMENT See <X14855> for complete genomic sequence.
Many of the URF's begin with GTG instead of ATG.
FEATURES
source
1. 4034
/organism="Thermoproteus tenax virus 1"
/strain="KRAL 10/12"
/db_xref="taxon:10479"
132..1556
/feature="viral protein TPX-VT3 (AA 1-474)"
/codon_start=1
/transl_table=11
/protein_id="CAA32838.1"
/db_xref="GI:584349"
/db_xref="SWISS-PROT:P19275"
/translation="MSYQITNNNTSNVQNIIVADIPLNLSLGENVFSQSGLSL
IYDGVWIKRSLPSQGITITASSGPNIDPTIALYNNSSYSNLIAGLVPTVS
IVDFGGAISAYAGDFLIVASPTGTPSSRLVLDVDRWATPTSLDVLGLSTANV
NDWFGVRYVAGQNSIEQKISGTSYVNEIDISQFAATDPLVMYLSINGTANY
KVYKQSNIGTVSGNYSTTPYGNPMAGYGVDKHYANFVLPYEDPQVTPISPP
SPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
TPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
TPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
complement(653..811)
/feature="unidentified reading frame (AA 1-52)"
/codon_start=1
/transl_table=11
/protein_id="CAA32839.1"
/db_xref="GI:62193"
/db_xref="SWISS-PROT:P19306"
/translation="MFEPLCYTTFVALPFDIKYITNGSYNANCDSISFTTEYVPD
IFCSMLTF"
complement(890..1243)
/feature="unidentified reading frame (AA 1-117)"
/codon_start=1
/transl_table=11
/protein_id="CAA32840.1"
/db_xref="GI:584350"
/db_xref="SPTREMBL:O88913"
/translation="MISYGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
VGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
VTCGSGSGNITKLA"
1991..2311
/feature="unidentified reading frame (AA 1-106)"

```

```

/codon_start=1
/transl_table=11
/protein_id="CAA32841.1"
/db_xref="GI:62195"
/db_xref="SWISS-PROT:P19307"
/translation="MIAEQIFILVLTLVHTAERIAKRSLSYSNVNVTYQSISFL
VGYTIVIMSHISHAVIISIMISVAYIVIKRNQQIQNDIEMFMFMGIAGLGI
EIVG"
complement(2265..2627)
/feature="unidentified reading frame (AA 1-120)"
/codon_start=1
/transl_table=11
/protein_id="CAA32842.1"
/db_xref="GI:584351"
/db_xref="SPTREMBL:O88914"
/translation="MGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
VGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
PIYFILLSLHLVQLSFQT"
2924..3386
/feature="unidentified reading frame (AA 1-61)"
/codon_start=1
/transl_table=11
/protein_id="CAA32843.1"
/db_xref="GI:584352"
/db_xref="SWISS-PROT:P19273"
/translation="MPSFQSTITRPSVGAIGPDIVYALYGYWNEQPLEIDKCSY
DLSTCSPVSSGPTIGDGYGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
GNSIKYIKIIFYNGYIYLITCDPGLLKCTNPINPSTCTQINAPTSFTVQEYFM
YTQGSYALLSYFQKNNKFLSLCNDGNTLYNINYSTTGILHYLHLSLEQYVE
II"
3401..3586
/feature="unidentified reading frame (AA 1-61)"
/codon_start=1
/transl_table=11
/protein_id="CAA32844.1"
/db_xref="GI:62198"
/db_xref="SWISS-PROT:P19273"
/translation="MYTQSYALLSYFQKNNKFLSLCNDGNTLYNINYSTTGIL
IHYLHLSLEQYVEIT"
BASE COUNT 1324 a 926 c 658 g 1126 t
ORIGIN

Query Match 2.1%; Score 47.8; DB 80; Length 4034;
Best Local Similarity 48.1%; Pred. No. 0.54;
Matches 136; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 1324 actatggtatcttagcaagttccatcttcgagggccgcaacgagagggcggtgctctggc 1383
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 942 ACTCTATATCTTCACCATCTCTCTACTCTACACGACACGCGCTACGCGCTACACCC 1001
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1384 tcaactacgatggcaccacgagaaattcgacggattctctcttcacacacgtca 1443
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1002 ACACCAAGCCCACTCGAGCGCTACTCTTACCCTACTCCGACACTACCGGATCC 1061
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1444 accaagctcccggtccaacggtggtgcogttgaaacttcttccaagacgctacctg 1503
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1062 ACTCCACACCTACTCCGACTCCGACACTACTCCAAACACCAACGCCAATCCGACACCG 1121
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1504 gtgctccctccatattcttaccattaccctctctgctgagggcaccacacacacacac 1563
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1122 ACTCCACGCGCAACGCGCTACTCCGACTCTTACCCTACTCCGACTCTTACCCTCC 1181
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1564 tgcacccacacggttaagctctctacggtgaggtctcttctgta 1606
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1182 ACTCCGACTCCACACCGACTCCGACACCAACGCGCTACGCGCTA 1224
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
LOCUS NPORPT2 749 bp DNA VRL 03-FEB-1999
DEFINITION Multicausid nuclear polyhedrosis virus of Orgyia
pseudotsugata (OpNPV), GC-rich repeated DNA from the SstI I

```







```

Db 481 GAGACGACGCTTCTTCGACACGCTGGTCCGCCGCT 518
||||| || ||||| ||||| || || || ||
RESULT 15
LOCUS OP075930 131993 bp DNA circular VRL 06-MAR-1998
DEFINITION Orgyia pseudotsugata nuclear polyhedrosis virus complete genome.
ACCESSION U75930
VERSION U75930.1 GI:2934903
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Ahrens,C.H., Russell,R.L., Funk,C.J., Evans,J.T., Harwood,S.H. and Rohrmann,G.F.
TITLE The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome
JOURNAL Virology 229 (2), 381-399 (1997)
MEDLINE 97271300
REFERENCE
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1998) Oregon State University, Agricultural Chemistry, Corvallis, OR 97331-7301, USA
AUTHORS
TITLE Submitted (06-MAR-1998) Oregon State University, Agricultural Chemistry, Corvallis, OR 97331-7301, USA
REMARK Sequence update by submitter replaced gi:1911246.
COMMENT On Mar 6, 1998 this sequence version replaced gi:1911246. similar to Autographa californica nuclear polyhedrosis virus (AcMNPV) complete genome: GenBank Accession Number U22858. Location/Qualifiers
1. 131993
/organism="Orgyia pseudotsugata nuclear polyhedrosis virus"
/db_xref="taxon:10450"
complement(123..947)
/notes="ORF1: p1-1, AcMNPV ORF10 homolog"
/codon_start=1
/product="protein kinase 1"
/protein_id="AAC59000.1"
/db_xref="GI:1911247"
/translation="MDAALQSLSRFVADCAVLAPKVVNGRFGKMDVLHRRPTTSKLEL
RTIAHSPSADLNLSDHSPVDMFYCYSSPTAWAIVMDYVPCPDLFETLQTO
GALDNALVNVROLCDALNDLHNGTGYTHNDVLENVLYFGARDRVYLCYGLCKRE
HSPVHGTLEFSPKIRRHNYARSDWYAGVYLAYKLLTGGRHPPERSADEVIDLAS
MRROQYNDPAALKNRLNWARDVFCLTRFNFCRSTDTYKAKHSFLASHNDYI"
946..2370
/notes="ORF2: AcMNPV ORF9 homolog"
/codon_start=1
/product="1629-capsid"
/protein_id="AAC59001.1"
/db_xref="GI:2934904"
/translation="MMERQYQSVRSYLNNOHNAALAGPELORVAGPEAHSVGRNVCDD
RAVNLNRQAVDLKLAEIDYADTATMQDQPEASSRHFFATLNRKLLLIGVQSPDQR
RNLNRLAEIALLRVDVNDVNAEVLGDFEYSKYISYQTTFAQTTPATDFFSRP
TSLRPQTSLSRQTSALPQOTFEDQPEMVSPSPSVHITPAIQQTPPATDFFSRP
SDSFVYVPGKERAVPDRFRKPPVPPKPEHLKSRSPSVATNAAGATVPAPPPPPFSAD
VTTSMPPPPPPPSADVTTSMPPPPPPVLDLATSMPPPPPPPPMVDLATSMPPPLINNA
INLLLDAMVAETNKAGDNRSALLDQIKOGTKLTKQPADGAPATDPRSTLLSEIRQ
GKTLKRLKIDQSTOTLLKVDYDTKTKTLKNTFYTNIDRISKQEQEKDRDLDTIT
KRPAVEHDTGNSGNSDPRDD"
complement(2533..3270)
/notes="ORF3: p1, AcMNPV ORF8 homolog"
/codon_start=1
/product="polyhedrin"
/protein_id="AAC59002.1"
/db_xref="GI:1911249"

/translation="MPDYSYRPTIGRTYVYDNKYKKNLGSVTKNAKRRKKULLHEDEE
KHLDPDHYVVAEDPFLGPKGNKQLTLFKEIRNVKPDITMKLILVNSGKREFLRTFRF
VEDSPFIYDOEVNDVFLVNNRPTPRNRCYKFLAQHRLDCDQVYVHEVIRIVERPSY
VGMNNEYRISLAKGGGCPIMNIHAEYTNCSFVNRVIRWENFYKPIVYITGDSSEEE
EILIEVSLVFKVEFAPDAPLFTGPAY"
complement(3350..3541)
/notes="ORF4"
/codon_start=1
/product="unknown"
/protein_id="AAC59003.1"
/db_xref="GI:1911250"
/translation="MRKNNGASQYPEHLTTTVHKRHLQLSCVWCDQDSMDPHSAKGL
MFEHMFCCRLIYVQOKIKH"
complement(3698..4093)
/notes="ORF5"
/codon_start=1
/product="unknown"
/protein_id="AAC59004.1"
/db_xref="GI:1911251"
/translation="MGASQNSQKILKLFYRWSQGTGAALDDEKDLHCLYDLERFVGA
HLNKRADDKARKKCAKALKRVEITAADRHMLEAAAPACADDGRWSTLSRAQLDDI
AREKEIVDRIHRLQLQDLSLLKQDRLLKR"
complement(4125..4739)
/notes="ORF6; left-2, AcMNPV ORF6 homolog"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAC59005.1"
/db_xref="GI:1911252"
/translation="MERYWNPAGIDGLKRSETYLDPHDFVGVLTLSPTVYVFERGLF
VMSGMRLALLAALPKQEPQPAVRFPQSRNRNVCLKACDAGSLAKVLAARVSNP
PNSKTMADLSAPRGNMYKRFENYLANVITCTCKTACLTGALLHFFYRMDKCV
GEVYHLLKAQDVKYPSNCAKMKVTKLCPQASMKCKGLNPICNF"
complement(4742..4975)
/notes="ORF7: AcMNPV ORF5 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59006.1"
/db_xref="GI:1911253"
/translation="MNRPTMRNTAAVTTDYDREQLRRELNSLRSSVHELCTRSTATGPD
CNRFLEAGDRAPAVIVKAAAGQGHSSLICDKV"
5020..5460
/notes="ORF8: AcMNPV ORF4 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59007.1"
/db_xref="GI:1911254"
/translation="MPGARVFRSLRTOEFKENVVAHVHMLGLRALIDGKVTSADEV
RRFGFSRNLVSVACMAVNVQVYAPDATIDMRHQPTTYIFRVQNCCHAMADVPAPDDHS
IARYLLAEGCAVLVIDHPLDVFGETEEGVNELLEYQRINAGGOL"
complement(5669..6151)
/notes="ORF9; ptp-2"
/codon_start=1
/product="protein tyrosine phosphatase 2"
/protein_id="AAC59008.1"
/db_xref="GI:1911255"
/translation="MYEASQIDHEFVGGYGDNDAMLRFIENHAIASVISLSDVA
PIRHALGPVGDHIHVYCEAAPTCALPNAMPALYDYVNRIGEGKRVLVHCYAGSR
SAALAVYILMRSQMPYQDALNAVQSKRRVAINDHFVRLATRCYRFVNDLAKPOIV"
complement(6129..6791)
/notes="ORF10; ptp-1, AcMNPV ORF1 homolog"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAC59009.1"
/db_xref="GI:1911256"
/translation="MLTSRWAVYITRTERSCNGCAWCTNGVRRRRQVHLPSVRARAM
FPDRHEVACCAVTEGTRLLCEKFLPNAELPEYVTSDEDRMTAAASVALRISALCAVI
DLNTARYDYGAQWVKGLLYKKIRVPGRAVDVDDIVAEIETVDEFFRCPTMLVAV
HWTGLNRSGLVLCVWVERLGVSPDTAIREFETARGHKIERTNYLQDLARKKHVRQO
PN"
6819..7814
/notes="ORF11: AcMNPV ORF11 homolog"
/codon_start=1

```

```

/product="unknown"
/protein_id="AAC59010.1"
/db_xref="GI:1911257"
/translation="MSLSKLLVAVYGYSYNLPDHYGESYHLYRIVVHEHLNTVYSN
ASCVRDRIATACRLNSGHLCPDVAOLLDVSEVAARI SAMFRCCGDAITGLCADMORALA
DIDRHAPLARVGRARRINFDALDADIPSDVTNNLGGIIGREMFHFRCSGLARVADVF
DPDIDAGDWMYHFCVLTVMELVACGAVPAGSATRLRDVAHAKHIGPNDEGNCAPALAA
VYGRFCAIGREHFAHHTKTHKALFLQFMRNDLTPADERHPCFVGKIDFGRCCKDTYTD
LKHADALYIHGTTTDQKNALFDLLCCCNASDIDADCYDCVNVNKFATQNKRYKM"
CDS
/notes="ORF12; AcMNPV ORF13 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59011.1"
/db_xref="GI:1911258"
/translation="MFSWHEGWNAADEQVNAEFDEQAYRYRAYVQDRAHSDLVLRDVF
RCHPFVFRYRLVDETRGCCSVDFCKGLIKSHOLLQRCNFRDQVROLNELVIGAP
PAQPSLGSFATKHLGLVOLQFQSFANKNEVLLAVGANKDHRDNLNKKIEVLNVH
KTLNNSDKFISAHKSFEKLVGEQFQFORLQTLDTKLNALQCAAPTARTAGVYFPR
DVTNPHLAVFMGRVEDRGVTQIAFARQGEHFRKRKLEFEEGMDVDRGRAPNPLLA
VHCIEEFGANGHKIRLRLPKRVIEVDCAVNVAKDIVKKAILNKT"
/notes="ORF13; lcf-1, AcMNPV ORF14 homolog"
/codon_start=1
/product="late expression factor 1"
/protein_id="AAC59012.1"
/db_xref="GI:1911259"
/translation="WAPCKYTPERVOMMDAIYNDSSRLAPMTDRPRVWHAHNFDS
AAQLFAYIVKNSISDVHVKPLEEGREWVIDADFKDCADKAELMLKLVNMGATAFMLFP
EGKEDAVQIFMSGNRGFHMLKFCGRKMDAPKSLREHFNVFQKPAKLVSGLDIRPG
SFADCVRRVAVHMYIGDAREDLVRYWFDVDRDVFVFCNANKQIRAPFSYNYKGGDYSRCL
TQQLQORIKACAGLAGGTPPTSK"
9893..11362
CDS
/notes="ORF14; egt, AcMNPV ORF15 homolog"
/codon_start=1
/product="ecdysteroid UDP-glucosyltransferase"
/protein_id="AAC59013.1"
/db_xref="GI:1911260"
/translation="MVFLIALTLATGARAASILAVLPTPAYSHVVYVAVVHALVK
NCHNVTVIPQLLDYAVODECGRVEOIDADMSAQYKLVASSGVFKEGVVADETIV
TADNYMGLIEMFKDFONANVRRLSTNRTFDVAVVEAPADYALVGHFLFRPAPVIOI
APGYGLAENFERERRAVABHPLHYETPGAALTRRGALSEWRNLNEFELLARRSDELL
KOQFGKSTPIRLQNDVOLLNHLHPVDDNNRPPVSVQVLGGGLHLAQALPQRLDA
PLERLNSVDGAVIVSESGIDTNSIAHEFLQMLLDTFANLNNTVLMKWDDVAAS
VALPRNLQAKWFSQTVLVRHKNVAVFVQAGLQSSDEALQARVPMVCLPMWMDQFHH
ARKLQGFVARALDTAAVSAPOQLAIREVIADEYARARIDKLRAVVEHDAAPDEKA
VKETERVIFENNVNVPARSLKTTAANMAYSDYFVRPEL"
11547..12140
CDS
/notes="ORF15; AcMNPV ORF16 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59014.1"
/db_xref="GI:1911261"
/translation="METAQPPISYAPPKRGAVCAVYRVVTVTTSVSDSGNNEDRLTQ
IVAQLRNLNFSKLQRRVRNMQKLIKKNKSVIANLAARLTQKTKHFAVTIR
KNVHTSGSEQFQRVLELCANGGEQVFCARADCARDRRRVAEALATAGAVVA
SAANKREFEIEDEELVSAKLIVQOVLHGDHSDT"
12109..12732
CDS
/notes="ORF16; AcMNPV ORF17 homolog"
/codon_start=1

```

```

Query Match      2.0%; Score 46; DB 80; Length 131993;
Best Local Similarity 44.3%; Pred. No. 1.6;
Matches 187; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

Oy 1323 aactatggtatcttagcaagtcattctcgagcccaacgagagaagcggtgctctg 1382
Db 75798 AACAAAGCCACTTTCCTCGCGAGCGCCGACCAACGCCATCGCGAGCGGACGCC 75739

Oy 1383 ctcaactacgatgcacacagagaattcgaaacgggattctcttcttcaactacgtc 1442
Db 75738 AACGCCATCGCGAGCGCGACCAACGCCGACGCCAACGCCGACGCCGACGCC 75679

```

```

Oy 1443 aaccagaacctcccggtcccaacggtggtgcccggttaaaaaacttcttcaagaagacccctaacctc 1502
Db 75678 ATCCCGCAGCGCGAGCGCGCGATATCAACAGCCGCGACGCATCGCGCGACGCTATCAACC 75619

Oy 1503 ggtgtctccctaaattctcttattacattaaacctctctgctgctgagccacacacacacacac 1562
Db 75618 AACGCCATCAACAGCGCGCGATCGCGAGCGGCATCGCGAGCGGCATCGCGAGCGGCATCGCGAGCGGC 75559

Oy 1563 atgacccacacgggtaacgctctctcaggtgaggtcttctgtaagcagagatctatgatgcc 1622
Db 75558 GACGCCATCGCGAGCGCGCATCAACAGCGCGCATCGCGAGCGGCATCGCGAGCGGCATCGCGAGCGGC 75499

Oy 1623 attgocaggggccctcagtgagggaagacacctgtcttcttattacctacgagagagaccggt 1682
Db 75498 GACGCCATCAACAGCGCGCGATCGCGAGCGGCATCGCGAGCGGCATCGCGAGCGGCATCGCGAGCGGC 75439

Oy 1683 ggttttagcagcactgttctctccctctcctcctcctcctcctcctcctcctcctcctcctcctc 1742
Db 75438 GACGCCATCGCGAGCGGCATCGCGAGCGGCATCGCGAGCGGCATCGCGAGCGGCATCGCGAGCGGC 75379

Oy 1743 ga 1744
Db 75378 GA 75377

```

Search completed: February 27, 2001, 00:55:28  
Job time: 15351 sec





Db 605 IKDLYYSERQSGPAMTYPVFAAAG-LLNHSSSQSYLYKSVLPYLRAAPAFQSEQ 663  
Qy 218 NDDDFLNGYSSKSI-FAEANEK-----GVSWLNVDGTNGEPEPDSLFFTYVNOTRS 269  
Db 664 SDONFLNGLTOPAFPLTANGFLOSLFGLTGIRY---SYEVDPTD---KKINRLLRF 717  
Qy 270 NVVPEFODAYLGLVP-----KFSYINPSCCGTNTNSMHTGNVSYGEVFAQIYDA 323  
Db 718 N--PIE-----LPLPGGTAIRNFKYMP----- 739  
Qy 324 IRQGFQWDKTLFTVDETFGGFYDH-----VPPPLAVRPDNLTYTETAKNGQXYT 373  
Db 740 -----GLDIIIDHNGCTIVHKSGDVPIHIKPNRSLIHQDDINFYNGSENERKPN 789  
Qy 374 LH--FDRLGMBTWPVSPYSKKGTYEQT 402  
Db 790 LERRDVDRGDPH-----RMDRYGT 809  
RESULT 4  
5223424-13  
; Patent No. 5223424  
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,  
; RICHARD D.  
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND  
; HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO  
; ACID SEQUENCE  
; NUMBER OF SEQUENCES: 16  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/225,032  
; FILING DATE: 27-JUL-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 78,519  
; FILING DATE: 27-JUL-1987  
; APPLICATION NUMBER: 933,107  
; FILING DATE: 20-NOV-1986  
; APPLICATION NUMBER: 902,887  
; FILING DATE: 02-SEP-1986  
; APPLICATION NUMBER: 887,140  
; FILING DATE: 17-JUL-1986  
; APPLICATION NUMBER: 823,102  
; FILING DATE: 27-JAN-1986  
; APPLICATION NUMBER: 773,430  
; FILING DATE: 06-SEP-1985  
; SEQ ID NO:13:  
; LENGTH: 579  
5223424-13  
Query Match 4.4%; Score 109.5; DB 5; Length 579;  
Best Local Similarity 21.4%; Pred. No. 0.042;  
Matches 86; Conservative 44; Mismatches 151; Indels 121; Gaps 22;  
Qy 110 FNDPDHVSVTGNL-----EFY-----GTYP-----NNGATASGKVVADOS 145  
Db 121 FNPADQQLISNMTEINLVSFEQEIFNVVLKTESATSPPSKIYNNDLTASLVALDTN 180  
Qy 146 GFLNAQLNDYKPLAPEATQVWGYTTEEVPTLVLDVDEFTTNSFCVPGGTPNRL 205  
Db 181 NTL-----PYTPAAPRSET---LGFY--PWLPT-----KPTQYRYLSCIRNLNP--- 221  
Qy 206 CALAGTAAGHCKDDDDFLNNGYSSKSI-FAEANEKGVSWLNVDG---TWG---EPEPDSLFF 260  
Db 222 -----TYTGQQQITDSIQTGLHSDIMEYTIENAVPIHLRTGDEFTSGIYHEDTKPLK 276  
Qy 261 TYVNOTSSRNVPVENFQ-----DAYLGLVPKFSYINPSCCGTNT-NSMHTGNVSYGE 314  
Db 277 TSWQINSLGLPKPVLTETPTGDOHRCGLP-----GANTRKGYHQTINNSYTE 326  
Qy 315 VFVKQYDAIROGP-QNDKTLFTVDETFGGFYDHVPPPLAVRPDNLTYTETAKNGQXYT 373  
Db 327 A-----TALRPAQVGYNTPMNFDYNGGPF-----TPIVPTADTQYDDERNGAIRET 376

Qy 374 L-----HFDRLGMBTWPVSPYSKKGTYE---BOYG-----TDPVTCKP 409  
Db 377 MGYOHGLHTSSQELERTFNPQSKCGRAPKQOQNOQAPLANLNTNNGTLLSDFPIGKGS 436  
Qy 410 -----APYSA---TSVLKTLGYLWDIE---DFTPRV 434  
Db 437 NKHEMNTLNTYGLPLTALNNTAPVFPNGQIWDKELDTDLKPL 478  
RESULT 5  
US-08-049-783-2  
; Sequence 2, Application US/08049783  
; Patent No. 5439881  
; GENERAL INFORMATION:  
; APPLICANT: Narva, Kenneth E  
; APPLICANT: Narva, George E  
; APPLICANT: Payne, Jewel M  
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active  
; TITLE OF INVENTION: Toxins Cloned from Bacillus thuringiensis Isolates  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jeff Lloyd  
; STREET: 2421 N.W. 41st Street  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/049,783  
; FILING DATE: 19930419  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lloyd, Jeff  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 2 (PS33F2):  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1257 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus thuringiensis  
; INDIVIDUAL ISOLATE: PS33F2  
; IMMEDIATE SOURCE:  
; CLONE: PS33F2a  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..1257  
US-08-049-783-2

Query Match 4.3%; Score 107; DB 1; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.25;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

Qy 6 LTGLGLASLG-LASPVTSYTSVREAPFGYKPGSKESIEINLKKVENIYVWLIEN--- 61  
Db 74 VLSVLTLPVEGTVASAASITVSFWPKIFGDKPNKAKNIFEELKPKQIEALLQDDITNYOD 133  
Qy 62 -----RSFNLGGVRRGLDNPINNGPCNKYKNASDPSSCKYCTQAKDYDSVND----- 112  
Db 134 AINQKFDLSLOKTINLTV--AIDNNDYVTAK-----TOLENLNLSILTSISIF 180





APPLICANT: Randall, John Brooks  
 APPLICANT: Meier, Henry  
 APPLICANT: Ulick, Heidi Jane  
 APPLICANT: Focerrada, Luis  
 APPLICANT: Schnepf, Harry E.  
 APPLICANT: Schwab, George E.  
 TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates  
 TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding  
 TITLE OF INVENTION: Hymenopteran-Active Toxins  
 NUMBER OF SEQUENCES: 39  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: David R. Saliwanchik  
 STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 32606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/304,626  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/887,980  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Saliwanchik, David R.  
 REGISTRATION NUMBER: 31,794  
 REFERENCE/DOCKET NUMBER: M/SCJ 104  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 904-375-8100  
 TELEFAX: 904-372-5800  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1257 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 STRANDEDNESS: single  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Bacillus thuringiensis  
 INDIVIDUAL ISOLATE: PS33F2  
 IMMEDIATE SOURCE:  
 CLONE: E. coli NM522(pMC2316) B-18785  
 FEATURE:  
 NAME/KEY: Protein  
 LOCATION: 1..1257  
 US-08-304-626-6

Query Match 4.3%; Score 107; DB 1; Length 1257;  
 Best Local Similarity 19.3%; Pred No. 0.25;  
 Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;  
 Qy 6 LITGGLASLG-LASPVTSVTSVREAPFGYKPGSKESINLKDKNVENVILLEN--- 61  
 Db 74 VLSVLTLPEVGTVAASASTVSTFWPKFEDGPNKAFELAPQIEALIQDITNYOD 133  
 Qy 62 ----RSEFNTLGGVRQGLDNPINNGPCNKNASDPSSGKYCTQAKDYDSVEND----- 112  
 Db 134 AINQKFFDSLOKTNILYTV--AIDNDYVYAK-----TQLENLSILTSISIF 180  
 Qy 113 -PDHVSVTGNIEFYTYTPNNGAIASGVV-ADQSGFLNAQNDYPKLAFEEATRVMGY 170  
 Db 181 IPEGVETG-GLPTYAMVANAHLRLDAIVNAEKLGFSDEVDTHKRY-----IKMTIHN 234  
 Qy 171 YTEEEVPTLDVDEFTTF--NSWFSCVPGPTNPRLCALAGTAAGHKNDDEFLNYGIS 228

Db 235 HTEAVIKAFNLGLDKFKSLDYSY-----NKKANY--- 264  
 Qy 229 SKSIFEANEKGVSWLNTDGTNGEFPDLSLFTTYN-----QTSRNVVPVENF 277  
 Db 265 IKGMTEKVLDLVALMPTEDPDHYQKEVEIEFTRTISSPIYQVPKKNQNTSSSIVPSDLF 324  
 Qy 278 FODAYLGVLPK--FS-----YINPSCCGT-----NTNSMHPGTG 309  
 Db 325 H---YQGLVKLEFSTRDNDGLAKIFTGIRNTEFYKSPNTHETVHVDVFSYNTQS---SSN 378  
 Qy 310 VSYG-----EVFKQIYD-----AIRQGQWQDKTLIF 336  
 Db 379 ISRGSSNPIDILNPIIISTCIRNSFYKAIAGSSVLVNFKDKGTQGYAFAAP----- 430  
 Qy 337 ITYDETGFGYDH--VPPPLAVRPDNLITYTETAKG-----OKYTLHFDRLGRMPTWVI 388  
 Db 431 -----TGWADHSFIESDGAPEGHKLNLYITSPGDTLRDFINVTLISTPTINELSTEKI 485  
 Qy 389 SPY-SKGYIEQYGTDPVTGKP 409  
 Db 486 KGFFAEKGYIKNOGIMKYIGKP 507  
 RESULT 8  
 US-08-316-301A-6  
 ; Sequence 6, Application US/08316301A  
 ; Patent No. 5753492  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schnepf, Harry E.  
 ; APPLICANT: Schwab, George E.  
 ; APPLICANT: Payne, Jewel M.  
 ; APPLICANT: Narva, Kenneth E.  
 ; APPLICANT: Focerrada, Luis  
 ; TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes  
 ; TITLE OF INVENTION: Which Code Therefor  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Saliwanchik & Saliwanchik  
 ; STREET: 2421 N.W. 41st Street, Suite A-1  
 ; CITY: Gainesville  
 ; STATE: FL  
 ; COUNTRY: USA  
 ; ZIP: 32606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/316,301A  
 ; FILING DATE: 30-SEP-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/871,510  
 ; FILING DATE: 23-APR-1992  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/693,018  
 ; FILING DATE: 03-MAY-1991  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/565,544  
 ; FILING DATE: 10-AUG-1990  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/084,653  
 ; FILING DATE: 12-AUG-1987  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/830,050  
 ; FILING DATE: 31-JAN-1992  
 ; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
REGISTRATION NUMBER: 35,589  
REFERENCE/DOCKET NUMBER: MA20CCCD1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6 (PS33F2):  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
CLONE: E. coli NM522(pMYC 2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-316-301A-6

Query Match 4.3%; Score 107; DB 1; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.25;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LITGLGLASLG-LASPVISEYTSVREAPFGYKPGSKESIEINKDKVENIWLLEN----61  
Db 74 VLSVLTLPETGVASASTVSIWPKIFGDKENAKNIFELKPKQIEALIQDDIITNYOD 133  
QY 62 ----RSFNLGGVRRGLDNPINNGPFCNYKNASDPSSGKYCTQAKDYDSVND----112  
Db 134 AINQKFDLSQKTLNLTV--AIDNDVVTAK-----TQLENLSILTSISIF 180  
QY 113 -PDHSTVGNLEFYCTYPNNGAIASGKVV-ADSGFLNAQLNDYPKLAPAEATROMGY 170  
Db 181 IPGVETG-GLPIYAVANNAHILLRDAIVNAEKLFGSDKEVDTHKKY-----IKRTIHN 234  
QY 171 YTEEVPTLVLDVDEFTF--NSWFSVCPGPTNPNRLCALAGTAGHGKNDDFLNYGIS 228  
Db 235 HTEAVIKAFGLDKFKSLDVNSY-----NKKANY---264  
QY 229 SKSITEAANEKGVSNLYDGTNGEPEPDSLFTYVN-----QTSRSNVVPVENF 277  
Db 265 IKGTEWLDLVALPFTFDHDYQVEIEFTTISPIYQVPKKNQNTSSSIVPSDLF 324  
QY 278 FQDYLGVLPK--FS-----YINPSCGT-----NTNSMHTGN 309  
Db 325 H---YQGLVLFLEFSTRDNDGLAKIFTGIRNTFYKSPNTHYHVDVFSYNTQS---SGN 378  
QY 310 VSYG-----EVEVKQIVD-----AIROGPQWDKTLF 336  
Db 379 ISRGSSNIPIDLNPISTCTRNFSFYKATAGSVLVNFKDGTQGYAFAPAP-----430  
QY 337 IYVDTGTGYDH--VPPPLAVRPDNLTYTAKNG-----QKTLHFDLGRGMPWTWY 388  
Db 431 ----TGAWDHSFIESDGAPEGHKLNIYVTPGDTLRDFINVYTLISTPTINELSTEKI 485  
QY 389 SPY-SKKGYIEQYTDPTVTKP 409  
Db 486 KGFPKGYIKNQIMKYGKP 507

## RESULT 9

US-08-611-928-6  
; Sequence 6, Application US/08611928  
; Patent No. 5824792  
; GENERAL INFORMATION:

APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Ulick, Heidi Jane  
APPLICANT: Foncerrada, Luis  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,928  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/158,232  
FILING DATE: 24-NOV-1993  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-611-928-6

Query Match 4.3%; Score 107; DB 2; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.25;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

```

Qy 6 LITGLLLASLG-LASPTVSEVTSREAPFGYKPGKSEIENLKDKVENIWLLEN--- 61
Db 74 VLSVLTLPVGTVASASTIVSFVFWPKIFGDKPNKKNIFEELKQIHALQOQDINYOD 133
Qy 62 ---RSFNILGVRROGLDNPINGPCNYKNASDPSSGKYCTQAKDYDSVFN----- 112
Db 134 AINOKKFDLSQKTINLYTV--AIDNNXYVTAK-----TOLENLSILTSISIF 180
Qy 113 -PDHSVTGNNLEFYGYTPNNGAIASGRV--ADOSGFLNALNDYKPLAPEATQVMGY 170
Db 181 IPEGYETG-GLPYAMVANAHILLRLDAIVNAEKLGSDEKVDTHKKY-----IKWTIHN 234
Qy 171 YTEEEVPLVDLVEFTTF--NSWFSVCVPGTPNPNRLCALAGTAAGHKNDLFLYNGIS 228
Db 235 HTEAVIKAFNLGLDKFSLDVNSY-----NKKANY--- 264
Qy 229 SKSIFEAAANEKGVSWLNYDGTNGEPEPDSLFETVYN-----QTSRSNVVPVNF 277
Db 265 IKGTEWLDLVALWPTFDPDHYQKEVEIEFTRTISSPIYQVPKNMONTSSSIVPSDLF 324
Qy 278 FQDAYLGLPK--FS-----YINPSCCGT-----NTNSMHTGN 309
Db 325 H---YOGDLVKLEFSTRDNDGLAKIFTGIRNTEFKSPNTHYHVDFTSYNTQS---SGN 378
Qy 310 VSYG-----EVEFKQIYD-----AIROGPQWDXTLF 336
Db 379 ISRGSSNPIDLNPISTCIRNSFYKAIGSSVLNFKDGTQGYAFAQAP----- 430
Qy 337 IYDETGGEYDH--VPPPLAVRPDLNLTETAKNG-----OKYTLHEDRIGGRMTWVI 388
Db 431 -----TGGAWDHSFIEDGAPCHKLNIYTPSGDTLRDFINVTTLSTPTINELSTEKI 485
Qy 389 SPY-SKKGYIEQYGTDPVTGKP 409
Db 486 KGFEAEKGYIKNQIMKYYGKP 507

```

## RESULT 10

```

US-09-173-891-6
: Sequence 6, Application US/09173891
: Patent No. 6077937
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Uick, Heidi Jane
: APPLICANT: Foncerrada, Luis
: APPLICANT: Schnepf, H. Ernest
: APPLICANT: Schwab, George E.
: APPLICANT: Fu, Jenny
: TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active
: TITLE OF INVENTION: Against Hymenopteran Pests
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/173,891
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/158,232
: FILING DATE:

```

```

: APPLICATION NUMBER: US 07/887,980
: FILING DATE: 22-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/797,645
: FILING DATE: 25-NOV-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/703,977
: FILING DATE: 22-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCJ104.C1
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1257 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: YES
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Bacillus thuringiensis
: INDIVIDUAL ISOLATE: FS33F2
: IMMEDIATE SOURCE:
: CLONE: E. coli NM522(pMYC2316) B-18785
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..1257
: US-09-173-891-6

```

Query Match 4.3%; Score 107; DB 3; Length 1257;

Best Local Similarity 19.3%; Pred. No. 0.25;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

```

Qy 6 LITGLLLASLG-LASPTVSEVTSREAPFGYKPGKSEIENLKDKVENIWLLEN--- 61
Db 74 VLSVLTLPVGTVASASTIVSFVFWPKIFGDKPNKKNIFEELKQIHALQOQDINYOD 133
Qy 62 ---RSFNILGVRROGLDNPINGPCNYKNASDPSSGKYCTQAKDYDSVFN----- 112
Db 134 AINOKKFDLSQKTINLYTV--AIDNNXYVTAK-----TOLENLSILTSISIF 180
Qy 113 -PDHSVTGNNLEFYGYTPNNGAIASGRV--ADOSGFLNALNDYKPLAPEATQVMGY 170
Db 181 IPEGYETG-GLPYAMVANAHILLRLDAIVNAEKLGSDEKVDTHKKY-----IKWTIHN 234
Qy 171 YTEEEVPLVDLVEFTTF--NSWFSVCVPGTPNPNRLCALAGTAAGHKNDLFLYNGIS 228
Db 235 HTEAVIKAFNLGLDKFSLDVNSY-----NKKANY--- 264
Qy 229 SKSIFEAAANEKGVSWLNYDGTNGEPEPDSLFETVYN-----QTSRSNVVPVNF 277
Db 265 IKGTEWLDLVALWPTFDPDHYQKEVEIEFTRTISSPIYQVPKNMONTSSSIVPSDLF 324
Qy 278 FQDAYLGLPK--FS-----YINPSCCGT-----NTNSMHTGN 309
Db 325 H---YOGDLVKLEFSTRDNDGLAKIFTGIRNTEFKSPNTHYHVDFTSYNTQS---SGN 378
Qy 310 VSYG-----EVEFKQIYD-----AIROGPQWDXTLF 336
Db 379 ISRGSSNPIDLNPISTCIRNSFYKAIGSSVLNFKDGTQGYAFAQAP----- 430
Qy 337 IYDETGGEYDH--VPPPLAVRPDLNLTETAKNG-----OKYTLHEDRIGGRMTWVI 388
Db 431 -----TGGAWDHSFIEDGAPCHKLNIYTPSGDTLRDFINVTTLSTPTINELSTEKI 485
Qy 389 SPY-SKKGYIEQYGTDPVTGKP 409

```

Db 486 KGPPAEKGYIKNOGIMKYGKP 507

# RESULT 11

US-09-076-137-6

; Sequence 6, Application US/09076137B

; Patent No. 6166195

; GENERAL INFORMATION:

; APPLICANT: Schnepf, Harry E.

; APPLICANT: Schwab, George E.

; APPLICANT: Payne, Jewel M.

; APPLICANT: Narva, Kenneth E.

; APPLICANT: Foncarrada, Luis

; TITLE OF INVENTION: No. 6166195el Nematode-Active Toxins and Genes Which Code

; TITLE OF INVENTION: Therefor

; FILE REFERENCE: MA-20CCCD2

; CURRENT APPLICATION NUMBER: US/09/076,137B

; CURRENT FILING DATE: 1998-05-12

; EARLIER APPLICATION NUMBER: 08/316,301

; EARLIER FILING DATE: 1994-09-30

; NUMBER OF SEQ ID NOS: 42

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 1257

; TYPE: PRT

; ORGANISM: Bacillus thuringiensis

US-09-076-137-6

Query Match 4.3%; Score 107; DB 3; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.25; Indels 166; Gaps 23;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LLTGLGLASLG-LASPVTSYTSVREAPFGYKPGKSEIENLKDKVENIVMLILEN---61

Db 74 VLSVLTLPVEGTVAASASTIVSFIMPKIFGDKPKNAKNIPEELKQIEALIQDITNYQD 133

QY 62 ----RSFNLILGVRROGLDNPINNGPCFNKYNASDPSSGKYCTQAKYDVSVDN----112

Db 134 AINQKFDLSLQKTINLTV--AIDNDYVTA-----TQLENLSILTSISIF 180

QY 113 -PDHVSVTGNLEFYTYTPNNGAIASGVV-ADSGFLNALQNDYKPKLAPEEATQVMGY 170

Db 181 IPEGYETG-GLPYAMVANAHILLRDAIVNAEKLGFSDKEVDTHKKY-----IKMTIHN 234

QY 171 YTEEVPVTLVDLVDDEFTT--NSWFSVCPGTPNRLCALAGTAAGHGKNDDDFLNYGIS 228

Db 235 HTEAVIKAFNLGLDKFKSLDVSRY-----NKKANY---264

QY 229 SKSIFPAANEKGVSWLNYDGTNGEPEPDSLFTTVN-----QTSRSNVVPPVENF 277

Db 265 IKGTEMVLDLVALMPTDPDHYQKEVEIEETRTISSPIYQVPVPMQNTSSIVPSDLF 324

QY 278 FQDAYLGVLPK--FS-----YINPSCGT-----NTNSMHPGTGN 309

Db 325 H---YQGLVKLEFSTRDNDGLAKIFTGIRNTFYKSPNTHETIYHVDFSYNTQS---SGN 378

QY 310 VSVG-----EVVKQIYD-----ATROGPQWDKTLFL 336

Db 379 ISKGSNPIPIDLNNPIITCIRNSFYKAIAGSSVLNFKDGTQGYAFAQAP-----430

QY 337 ITVDEYGGFYDH--VPPPLAVRPDNLITYETAKNG-----QKYLHFDRLGGRMPTWI 388

Db 431 ----TGGAWDSHFTESDGAPEGHKLNLYITSPGDTLRDFINVTYLISTPTINELSTEKI 485

QY 389 SPY-SKKGIEQYGDVPTGKP 409

Db 486 KGPPAEKGYIKNOGIMKYGKP 507

# RESULT 12

PCT-US92-03624-6

; Sequence 6, Application PC/TUS9203624

; GENERAL INFORMATION:  
; APPLICANT: Schnepf, Harry E.  
; APPLICANT: Schwab, George E.  
; APPLICANT: Payne, Jewel M.  
; APPLICANT: Narva, Kenneth E.  
; APPLICANT: Foncarrada, Luis  
; TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes  
; TITLE OF INVENTION: Which Code Therefor  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL USA  
; COUNTRY: USA  
; ZIP: 32606

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/03624  
; FILING DATE: 19920501  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA20C2C1C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 6 (PS33F2):  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1257 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus thuringiensis  
; INDIVIDUAL ISOLATE: PS33F2  
; IMMEDIATE SOURCE:  
; CLONE: E. coli NM522(pMC 2316) B-18785  
; FEATURE:  
; NAME/KEY: protein  
; LOCATION: 1..1257  
; PCT-US92-03624-6

Query Match 4.3%; Score 107; DB 4; Length 1257;  
Best Local Similarity 19.3%; Pred. No. 0.25; Indels 166; Gaps 23;  
Matches 97; Conservative 73; Mismatches 166; Indels 166; Gaps 23;

QY 6 LLTGLGLASLG-LASPVTSYTSVREAPFGYKPGKSEIENLKDKVENIVMLILEN---61

Db 74 VLSVLTLPVEGTVAASASTIVSFIMPKIFGDKPKNAKNIPEELKQIEALIQDITNYQD 133

QY 62 ----RSFNLILGVRROGLDNPINNGPCFNKYNASDPSSGKYCTQAKYDVSVDN----112

Db 134 AINQKFDLSLQKTINLTV--AIDNDYVTA-----TQLENLSILTSISIF 180

QY 113 -PDHVSVTGNLEFYTYTPNNGAIASGVV-ADSGFLNALQNDYKPKLAPEEATQVMGY 170

Db 181 IPEGYETG-GLPYAMVANAHILLRDAIVNAEKLGFSDKEVDTHKKY-----IKMTIHN 234

QY 171 YTEEVPVTLVDLVDDEFTT--NSWFSVCPGTPNRLCALAGTAAGHGKNDDDFLNYGIS 228

Db 235 HTEAVIKAFNLGLDKFKSLDVSRY-----NKKANY---264

QY 229 SKSIFPAANEKGVSWLNYDGTNGEPEPDSLFTTVN-----QTSRSNVVPPVENF 277

```

Db 265 IKGTEMLDLVALMPTDPPHYQKEVEIEFTRTISPIYQVPKNNQNTSSSIVPSDLF 324
QY 278 FQDAYLGLVLPK--FS-----YINPSCGT-----NTNSMHPGN 309
Db 325 H---YQGLVKLEFSTRDNDGLAKFTGIRNTEFYKSPNTHETVHVDVSYNTQS---SGN 378
QY 310 VSYG-----EVFKQIYD-----AIRQGFQWOKTLIF 336
Db 379 ISRGSSNPIDLNPIISTCIRNSFYKATAGSSVLVNFKDGTOGYAFQAAP----- 430
QY 337 ITYDETGGFYDH--VPPPLAVRPDNLFTYTETAKNG-----QKTYLHFDRLGGRMPTWVI 388
Db 431 ----TGGAMHDSFIESDGAPEGHKLANIVTSPGDLRDFINVTTLISTTINELSTEKI 485
QY 389 SPY-SKGGYIEQYGDPTVTKP 409
Db 486 KGFAEKGYIKNOGIMKYGKP 507

```

## RESULT 13

US-08-119-361-5

: Sequence 5, Application US/08119361

: Patent No. 5523390

: GENERAL INFORMATION:

: APPLICANT: Travis, James

: APPLICANT: Potempa, Jan

: APPLICANT: Barr, Philip J.

: APPLICANT: Pavloff, Nadine

: TITLE OF INVENTION: Porphyromonas gingivalis Arginine-specific Proteinase

: NUMBER OF SEQUENCES: 16

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Greenlee and Winner, P.C.

: STREET: 5370 Manhattan Circle, Suite 201

: CITY: Boulder

: STATE: CO

: COUNTRY: USA

: ZIP: 80303

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: OPERATING SYSTEM: IBM PC compatible

: SOFTWARE: PatentIn Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/119,361

: FILING DATE: 10-SEP-1993

: CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:

: NAME: Ferber, Donna M.

: REGISTRATION NUMBER: 33878

: REFERENCE/DOCKET NUMBER: 21-93

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 303-499-8080

: TELEFAX: 303-499-8089

: INFORMATION FOR SEQ ID NO: 5:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 737 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

US-08-119-361-5

```

Query Match      4.2%  Score 104.5; DB 1; Length 737;
Best Local Similarity 20.0%; Pred. No. 0.19;
Matches 104; Conservative 69; Mismatches 159; Indels 189; Gaps 30;

```

```

QY 36 YRPGSKESIEHLKDKVENIWLILNRSFNDILGCVRRGCLDPLNNGPCNKYNKASDPS 95
Db 223 YEPGRYTYVE-----EKONGRMIVIAKKYE-----DFVDMKNGRG-- 262
QY 96 SGKYCTQAKDYDSVFNDPOHSVTGNLFEY--GTYTPNNGAIASCKVVAOSGFLNAQLN 153

```

```

Db 263 ---LRTEVK---VAEDIASPTTANALQOQVKEVEKEGNDLTVLLVGDH-----K 307
QY 154 DYP-KLAPEATQVMG-----YYTE-----EVPITLVLDVDEF-----TTENS 191
Db 308 DIPAKITPGIKSDQYGOIVGNHDYNEVFTGRFSCSKEDLKTQIDRTIHYERNITTEDK 367
QY 192 WFSVCPGPTNPRLC---ALAGTAAGHGKNDQDFLNY-----GISS 229
Db 368 WLG-----QALCIASAEQGPSADNGESDIQHENVIANLLTQGYTKIKICYDPGVTP 419
QY 230 KSIFEANKEGWSLNYDGTNGEPEPDSLF--TYVNTQTSRSNVVVFENFQDA----- 281
Db 420 KNIIDAFN--GGISLVNNTYCHGSETAWGTSHFGTTHVKQLTNSNLP---FIEDVACVNGD 475
QY 282 YLGVLPKF-----SYNPS-----CCGTNTNSMHP 306
Db 476 FLFSMPCFAEALMRAQKDKPTGTVAIIASTINQSWASPMRGODEMNEILCEKHNNIKR 535
QY 307 T-GNVSYGEVY---KQYDAIROGPQW---DKYLLFITYDETTGGFYDHVPPPLAV-RP 357
Db 536 TFGVTNMGFMAMVERIKKDKGKMLDTWTVFGDPSLLVRTL-----VPTKMOVTPAP 586
QY 358 DNLYTYTET-----AKNGQY-----TLHFDRLGGRMP-TWVISPY 391
Db 587 AQINLTDSVNSCDYNGAIIATISANGKMEGSAVVENGATATINLTGLTNSTLTLTVVGY 646
QY 392 SKGGYIEQYGTDPVTKPAPY-----SATSVLKTLGLYLD 426
Db 647 NKEIVIKTIINT---GEPNPYPVSUNLTATQCGKVTLKWD 684

```

## RESULT 14

US-08-336-308A-4

: Sequence 4, Application US/08336308A

: Patent No. 6017532

: GENERAL INFORMATION:

: APPLICANT: Travis, James

: APPLICANT: Potempa, Jan S.

: APPLICANT: Barr, Philip J.

: APPLICANT: Pavloff, Nadine

: TITLE OF INVENTION: Porphyromonas gingivalis

: NUMBER OF SEQUENCES: 16

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

: STREET: 5370 Manhattan Circle, Suite 201

: CITY: Boulder

: STATE: Colorado

: COUNTRY: US

: ZIP: 80303

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: OPERATING SYSTEM: IBM PC compatible

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/336,308A

: FILING DATE: 08-NOV-1994

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 08/119,361

: FILING DATE: 10-SEP-1993

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 08/265,441

: FILING DATE: 24-JUN-1994

: ATTORNEY/AGENT INFORMATION:

: NAME: Ferber, Donna M.

: REGISTRATION NUMBER: 33,878

: REFERENCE/DOCKET NUMBER: 21-93C

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (303) 499-8080

: TELEFAX: (303) 499-8089

: INFORMATION FOR SEQ ID NO: 4:

```

; SEQUENCE CHARACTERISTICS:
;   LENGTH: 737 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-336-308A-4

```

```

Query Match
Best Local Similarity 4.2%; Score 104.5; DB 3; Length 737;
Matches 104; Conservative 69; Mismatches 159; Indels 189; Gaps 30;

QY 36 YKPGSKESIEHLKDKVENIVWLLLENRSFDNLLGGVRRQGLDNPINNGPFCNKKNSADSPS 95
Db 223 YEGRTPTVE-----EKQNGRMIVIVAKYIE---GDIK-----DFVDMKNQRG-- 262

QY 96 SGKYCTQAKDYSDVFNPDHVSVTGNLFY--GYTPNNGAIGASGVVADQSGFLNAQLN 153
Db 263 ---LRTEVK---VAEDIASPVANAIOQFVKQYEKEGNDLTYVLLVGDH-----K 307

QY 154 DYP-KLAPEATQVMG-----YYTE-----BEVPTLVDLDEF-----TTFNS 191
Db 308 DIPAKITPGIKSDQVYQIVGNDHNEVFIGRFSCSKEDLKTQIDRTIHYERNITTEBK 367

QY 192 WFCSCVPGTNPRLC---ALAGTAAGHKNDLDFNY-----GISS 229
Db 368 WLG-----QALCIASAEGPSADNGESDIQHENVIANLLTQYGYTKIKCYDPGVTP 419

QY 230 KSIFEAAKEGVSWLNYDGTNGEPEFDSLFF--TYVNOTSRNVVPPVENFFQDA----- 281
Db 420 KNIIDAFN--GGISLVNYTGHGSETAWGTSHGTHVKQLTNSQLP---FFFDVACVNGD 475

QY 282 YLGVLPKF-----SYNPS-----CCGTNTNSMHP 306
Db 476 FLSPMPCFAELMRAQKDGKPTGTVAIIASTINQSWASPMRGQDEMNEILCEKHPNNIKR 535

QY 307 T-GNVSYGEVFV---KQIYDAIROGPW---DKTLFFITYDETGGFYDHPPLAV-PP 357
Db 536 TFGGVTMNGMEAMVEKYKDGKMLDTWTFVGDPSLLVRTL-----VPTKMQVTAP 586

QY 358 DNLTYTET-----AKNGQY-----TLHFDRLGGMP-TWVISPY 391
Db 587 AQINLUDASVNVSCDYNAGAIATISANGKMGSAVNGTATINLTGLTNESTLTITLVVGY 646

QY 392 SKGYIEQYDTPVTGKPAPY-----SATSVLKTIGYLWD 426
Db 647 NKETVIKTINTN---GEPNPYQPVSNLTATITQGVKTLKWD 684

```

## RESULT 15

```

US-08-822-324-4
; Sequence 4, Application US/08822324
; Patent No. 6129917
; GENERAL INFORMATION:
; APPLICANT: Potempa, Jan S.
; APPLICANT: Travis, James
; APPLICANT: Genco, Caroline A.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,945
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 103-95 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 488-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-822-324-4

```

```

Query Match
Best Local Similarity 4.2%; Score 104.5; DB 3; Length 737;
Matches 104; Conservative 69; Mismatches 159; Indels 189; Gaps 30;

```

```

QY 36 YKPGSKESIEHLKDKVENIVWLLLENRSFDNLLGGVRRQGLDNPINNGPFCNKKNSADSPS 95
Db 223 YEGRTPTVE-----EKQNGRMIVIVAKYIE---GDIK-----DFVDMKNQRG-- 262

QY 96 SGKYCTQAKDYSDVFNPDHVSVTGNLFY--GYTPNNGAIGASGVVADQSGFLNAQLN 153
Db 263 ---LRTEVK---VAEDIASPVANAIOQFVKQYEKEGNDLTYVLLVGDH-----K 307

QY 154 DYP-KLAPEATQVMG-----YYTE-----BEVPTLVDLDEF-----TTFNS 191
Db 308 DIPAKITPGIKSDQVYQIVGNDHNEVFIGRFSCSKEDLKTQIDRTIHYERNITTEBK 367

QY 192 WFCSCVPGTNPRLC---ALAGTAAGHKNDLDFNY-----GISS 229
Db 368 WLG-----QALCIASAEGPSADNGESDIQHENVIANLLTQYGYTKIKCYDPGVTP 419

QY 230 KSIFEAAKEGVSWLNYDGTNGEPEFDSLFF--TYVNOTSRNVVPPVENFFQDA----- 281
Db 420 KNIIDAFN--GGISLVNYTGHGSETAWGTSHGTHVKQLTNSQLP---FFFDVACVNGD 475

QY 282 YLGVLPKF-----SYNPS-----CCGTNTNSMHP 306
Db 476 FLSPMPCFAELMRAQKDGKPTGTVAIIASTINQSWASPMRGQDEMNEILCEKHPNNIKR 535

QY 307 T-GNVSYGEVFV---KQIYDAIROGPW---DKTLFFITYDETGGFYDHPPLAV-PP 357
Db 536 TFGGVTMNGMEAMVEKYKDGKMLDTWTFVGDPSLLVRTL-----VPTKMQVTAP 586

QY 358 DNLTYTET-----AKNGQY-----TLHFDRLGGMP-TWVISPY 391
Db 587 AQINLUDASVNVSCDYNAGAIATISANGKMGSAVNGTATINLTGLTNESTLTITLVVGY 646

QY 392 SKGYIEQYDTPVTGKPAPY-----SATSVLKTIGYLWD 426
Db 647 NKETVIKTINTN---GEPNPYQPVSNLTATITQGVKTLKWD 684

```

Search completed: February 26, 2001, 16:20:37  
Job time: 43367 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2001, 02:55:27 ; Search time 126.87 Seconds  
(without alignments)  
2944.507 Million cell updates/sec

Title: US-09-426-072-1

Perfect score: 2318

Sequence: 1 aattctggacacaaatccc.....gcaccccgatataattaatt 2318

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2318	100.0	2318	3	US-09-426-072-1
2	52.6	2.3	7218	1	US-08-232-463-14
3	46.4	2.0	925	3	US-08-858-003-1
4	46.4	2.0	925	3	US-09-078-166-1
5	45.4	2.0	5511	3	US-08-928-361B-2
6	45.4	2.0	7334	3	US-08-928-361B-1
7	45.2	1.9	1425	2	US-08-578-551-2
8	45.2	1.9	1425	3	US-09-190-982-2
9	44.8	1.9	5163	3	US-08-700-651-1
10	44.8	1.9	5163	3	US-08-928-361B-4
11	44.8	1.9	5318	3	US-08-700-651-2
12	44.8	1.9	5318	3	US-08-928-361B-3
13	42.8	1.8	28804	2	US-08-592-874-1
14	42.8	1.8	28804	3	US-09-096-942-2
15	42.8	1.8	28804	3	US-09-096-867-2
16	39	1.7	1155	2	US-08-387-942C-7
17	38.8	1.7	1227	3	US-09-074-912-3
18	38.8	1.7	2064	1	US-08-343-428-1
19	38.6	1.7	36519	3	US-08-923-137-2
20	38.4	1.7	1931	2	US-09-130-114-2
21	38	1.6	966	2	US-08-766-738-2
22	37.8	1.6	2093	1	US-08-287-001A-1
23	37.8	1.6	2093	4	PCT-US95-09941-1
24	37.4	1.6	1155	2	US-08-387-942C-18
25	37.4	1.6	12588	2	US-08-387-942C-1
26	37	1.6	3300	1	US-08-194-290-6
27	37	1.6	3300	2	US-08-614-377A-6
28	36.8	1.6	1531	1	US-08-449-986-1

29	36.8	1.6	1531	2	US-08-756-855-1	Sequence 1, Appli
30	36.6	1.6	2214	3	US-08-864-038A-1	Sequence 1, Appli
31	36.6	1.6	3331	3	US-08-864-038A-2	Sequence 2, Appli
32	36.6	1.6	3331	3	US-08-864-038A-4	Sequence 4, Appli
33	36.4	1.6	1284	3	US-08-911-894-72	Sequence 72, Appli
34	36.2	1.6	309	1	US-08-209-747-3	Sequence 3, Appli
35	36.2	1.6	309	1	US-08-458-298-3	Sequence 3, Appli
36	36.2	1.6	1124	2	US-08-578-551-1	Sequence 1, Appli
37	36.2	1.6	1124	2	US-09-190-982-1	Sequence 1, Appli
38	36	1.6	458	3	US-09-141-000-4	Sequence 4, Appli
39	36	1.6	624	2	US-08-426-599B-3	Sequence 3, Appli
40	36	1.6	1182	1	US-07-640-476-4	Sequence 4, Appli
41	36	1.6	1185	5	5290690-1	Patent No. 5290690
42	36	1.6	1566	5	5290690-3	Patent No. 5290690
43	35.6	1.5	8931	3	US-09-028-934-28	Sequence 28, Appli
44	34.8	1.5	38506	3	US-09-320-878-19	Sequence 19, Appli
45	34.4	1.5	2061	1	US-08-204-656B-9	Sequence 9, Appli

## ALIGNMENTS

RESULT 1  
US-09-426-072-1  
; Sequence 1, Application US/09426072  
; Patent No. 6146869  
; GENERAL INFORMATION:  
; APPLICANT: Kimberly M. Brown  
; TITLE OF INVENTION: Polypeptides Having Phospholipase B  
; FILE REFERENCE: 5951.000-US  
; CURRENT APPLICATION NUMBER: US/09/426.072  
; CURRENT FILING DATE: 1999-10-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 2318  
; TYPE: DNA  
; ORGANISM: Aspergillus oryzae  
US-09-426-072-1

Query Match	100.0%	Score 2318	DB 3	Length 2318
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2318	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
Qy	1	aattctggacacaaatcccctttgagagtgatagtaaggaggaggaactgagatgctaag	60	
Db	1	aattctggacacaaatcccctttgagagtgatagtaaggaggaggaactgagatgctaag	60	
Qy	61	ctcaagttctgccttgaggtaggctgcccattgtcttggtttaggtcttctccacc	120	
Db	61	ctcaagttctgccttgaggtaggctgcccattgtcttggtttaggtcttctccacc	120	
Qy	121	ctagtagttagccattgcttcccaatcccatcccatcccatcccatcccatcccat	180	
Db	121	ctagtagttagccattgcttcccaatcccatcccatcccatcccatcccatcccat	180	
Qy	181	gaccaaggtgtcaaatagagagcccaactatgctgcctccatctcatattttttt	240	
Db	181	gaccaaggtgtcaaatagagagcccaactatgctgcctccatctcatattttttt	240	
Qy	241	gtcgataagactgtcatccctctccgggtggtctccacgaagaacgatttgtaagatgg	300	
Db	241	gtcgataagactgtcatccctctccgggtggtctccacgaagaacgatttgtaagatgg	300	
Qy	301	atagcattaggtgccttgcctgtatataaagcccaagctccctgcctcgatgatat	360	
Db	301	atagcattaggtgccttgcctgtatataaagcccaagctccctgcctcgatgatat	360	
Qy	361	gcgcttaacccgcagcacgacgctggtatccgaatgggacggaaggtagctctc	420	
Db	361	gcgcttaacccgcagcacgacgctggtatccgaatgggacggaaggtagctctc	420	

Db 361 gccttcaaacccgacgcagcgcacgtggtatccgaaatgggacggaaaagttagctcttc 420  
Qy 421 cagtggtgtatataagcctttccctggccgggtccagactccctagcaagcagcattgtt 480  
Db 421 cagtggtgtatataagcctttccctggccgggtccagactccctagcaagcagcattgtt 480  
Qy 481 caagctcagcactcaaatccgcagagtaaaatgaatcaacgcctctgcttactgtctc 540  
Db 481 caagctcagcactcaaatccgcagagtaaaatgaatcaacgcctctgcttactgtctc 540  
Qy 541 ggcctcttgccctctctcgttctgttagcctgtcaagctccagagtagatagcagcgtgga 600  
Db 541 ggcctcttgccctctctcgttctgttagcctgtcaagctccagagtagatagcagcgtgga 600  
Qy 601 gaagccctcttcagatacaagcctgtctcaagagtagcattgagaaacttgaaggaaga 660  
Db 601 gaagccctcttcagatacaagcctgtctcaagagtagcattgagaaacttgaaggaaga 660  
Qy 661 gtccgaacattgtctggtcttattctcgagaaacaggttctgtgcccctacgatatctat 720  
Db 661 gtccgaacattgtctggtcttattctcgagaaacaggttctgtgcccctacgatatctat 720  
Qy 721 atgtggaataataatctcaattcagctgtgtgtatgaagacagcacttaacatat 780  
Db 721 atgtggaataataatctcaattcagctgtgtgtatgaagacagcacttaacatat 780  
Qy 781 ggtccagactcttcgataacattctggtgagcgtggtgcccgaagagactggacaacccga 840  
Db 781 ggtccagactcttcgataacattctggtgagcgtggtgcccgaagagactggacaacccga 840  
Qy 841 tcaacaaacgcccgtcttcgataacattctggtgagcgtggtgcccgaagagactggacaacccga 900  
Db 841 tcaacaaacgcccgtcttcgataacattctggtgagcgtggtgcccgaagagactggacaacccga 900  
Qy 901 gtaactcagcagcagactatgattccgtgttcaagatccagacactccgtgactggtta 960  
Db 901 gtaactcagcagcagactatgattccgtgttcaagatccagacactccgtgactggtta 960  
Qy 961 ataaacttgagttctacggaacttacaccccaaaacaaatggtgcgattgcccagtgcaagg 1020  
Db 961 ataaacttgagttctacggaacttacaccccaaaacaaatggtgcgattgcccagtgcaagg 1020  
Qy 1021 tgcgtccgacacagctggctcttcacacagcagcttaacgactaaccacaaactggccc 1080  
Db 1021 tgcgtccgacacagctggctcttcacacagcagcttaacgactaaccacaaactggccc 1080  
Qy 1081 cagaagaggcgaacagcaagtgatgggatactatacggagagagaggttctcagctcg 1140  
Db 1081 cagaagaggcgaacagcaagtgatgggatactatacggagagagaggttctcagctcg 1140  
Qy 1141 tggaccttgagtagtctactcttcaacagctgttctcgtgtgttctgttctggtgta 1200  
Db 1141 tggaccttgagtagtctactcttcaacagctgttctcgtgtgttctgttctggtgta 1200  
Qy 1201 gtcgataaaatctgtatatttttaatttgactaacgcttcaaaagccttcaaaacccca 1260  
Db 1201 gtcgataaaatctgtatatttttaatttgactaacgcttcaaaagccttcaaaacccca 1260  
Qy 1261 accgcttgctcctcgcaggaacagcgtgctggcagatggcaagatgacgatgactcc 1320  
Db 1261 accgcttgctcctcgcaggaacagcgtgctggcagatggcaagatgacgatgactcc 1320  
Qy 1321 tgaactatggtatcttagcagagtcacatcttcgagggccgcgaacgagagggcgtgtct 1380  
Db 1321 tgaactatggtatcttagcagagtcacatcttcgagggccgcgaacgagagggcgtgtct 1380  
Qy 1381 ggcctcaactacagtgccacgaagagagatccgaacggattctcttcttccactacg 1440  
Db 1381 ggcctcaactacagtgccacgaagagagatccgaacggattctcttcttccactacg 1440  
Qy 1441 tcaaccagacccctccggttccaaogtgggtgcccgttggaaaacttcttccaaagcgtacc 1500  
Db 1441 tcaaccagacccctccggttccaaogtgggtgcccgttggaaaacttcttccaaagcgtacc 1500

Qy 1501 tcgggtgctctccctaaattctcttacattaaacccctctctgtgctgagcacaacacaact 1560  
Db 1501 tcgggtgctctccctaaattctcttacattaaacccctctctgtgctgagcacaacacaact 1560  
Qy 1561 ccattgccccccacccggttaacgtctctacggtgaggtcttcgtcaagcagatcatgatg 1620  
Db 1561 ccattgccccccacccggttaacgtctctacggtgaggtcttcgtcaagcagatcatgatg 1620  
Qy 1621 ccattgccccccacccggttaacgtcttcgaggaagacccctgctcttattacacacgagacg 1680  
Db 1621 ccattgccccccacccggttaacgtcttcgaggaagacccctgctcttattacacacgagacg 1680  
Qy 1681 gtggcttctacgacctgtccctcccccctctgcccctgcccgcggaacacacacacacac 1740  
Db 1681 gtggcttctacgacctgtccctcccccctctgcccctgcccgcggaacacacacacacac 1740  
Qy 1741 ctgagactgcgaagacaggtcagaaatcacactcttcacacttcacacttcacacttcac 1800  
Db 1741 ctgagactgcgaagacaggtcagaaatcacactcttcacacttcacacttcacacttcac 1800  
Qy 1801 tgcgcagactgggttctccctctcacagtaagaagggatatactgcagcagtagcagaaacgg 1860  
Db 1801 tgcgcagactgggttctccctctcacagtaagaagggatatactgcagcagtagcagaaacgg 1860  
Qy 1861 atcccgctcacgggcaagcccctccctacagtgctacactccgctcctcaagactctcgat 1920  
Db 1861 atcccgctcacgggcaagcccctccctacagtgctacactccgctcctcaagactctcgat 1920  
Qy 1921 atctctggagacatcagagacttcccccctctgctgcccacactctccatcttctcagacc 1980  
Db 1921 atctctggagacatcagagacttcccccctctgctgcccacactctccatcttctcagacc 1980  
Qy 1981 tgcgcgacgacttgcgtgaggtgctctctctctctctctctctctctctctctctctct 2040  
Db 1981 tgcgcgacgacttgcgtgaggtgctctctctctctctctctctctctctctctctctct 2040  
Qy 2041 cgggtatagctcagtcgctgcagcagcagaggtataaagtgatgatgaaacgttctgac 2100  
Db 2041 cgggtatagctcagtcgctgcagcagcagaggtataaagtgatgatgaaacgttctgac 2100  
Qy 2101 ttcagtgaaacgattacattgtttatagagcattgtttgtcttatagctacgcttagagcag 2160  
Db 2101 ttcagtgaaacgattacattgtttatagagcattgtttgtcttatagctacgcttagagcag 2160  
Qy 2161 cgcgagatgataaataaaagctgggtttctctctctctctctctctctctctctctctct 2220  
Db 2161 cgcgagatgataaataaaagctgggtttctctctctctctctctctctctctctctctct 2220  
Qy 2221 tttgataacatggggttttaaaaggaatgttttcttctctctctctctctctctctctct 2280  
Db 2221 tttgataacatggggttttaaaaggaatgttttcttctctctctctctctctctctctct 2280  
Qy 2281 cgttaactttacgactgaccccccggatatacattatt 2318  
Db 2281 cgttaactttacgactgaccccccggatatacattatt 2318

## RESULT 2

US-08-232-463-14

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley &amp; Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA





RESULT 4  
US-09-078-166-1  
; Sequence 1, Application US/09078166  
; Patent No. 6063561  
; GENERAL INFORMATION:  
; APPLICANT: Katz, Leonard  
; APPLICANT: Stassi, Diane L.  
; APPLICANT: Summers Jr., Richard G.  
; APPLICANT: Ruan, Xiaolan  
; APPLICANT: Pereda-Lopez, Ana  
; APPLICANT: Kakavas, Stephan J.  
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES  
; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Rd.  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/078,166  
; FILING DATE: 16-MAY-1979  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dianne Casuto  
; REGISTRATION NUMBER: P-40,943  
; REFERENCE/DOCKET NUMBER: 4952.US.P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (847)-938-3137  
; TELEFAX: (847)-938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 925 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-078-166-1

Query Match 2.0%; Score 46.4; DB 3; Length 925;  
Best Local Similarity 44.6%; Pred. No. 0.00042;  
Matches 182; Conservative 0; Mismatches 226; Indels 0; Gaps 0;  
QY 1374 gtgtcttggtctcaactcgtatggcaccacgagaaatcgaacggattctctctcttc 1433  
DB 259 GTCTGGCGGACACTCGGTGGGAGAAATCACCGCAGCCACGCTCTCCGGCACCTCACCCCTC 318  
QY 1434 acctacgtcaaccagacctccgggtcccaacgtggtgccegtgaaactcttccaagac 1493  
DB 319 CCGGAGCGCGTCCACCTGTCACACACCGCGGACCCCTCATGCAAAACCTGCCCCCGGC 378  
QY 1494 gcttacctggtgtctccctcaaatctcttattacacctctctgtgcygaccacac 1553  
DB 379 GGGCGCATGCTGGCGGTGCGCACCGACCCGACACCTCCACACCCACCTCGCAACACAC 438  
QY 1554 accaactccatgcacccaccggtacgtctctacagtgaggttttgtcaagcagatc 1613  
DB 439 CAGCAGACCATCTCCATCGCGGCGCATCAACGGCCCCCAGCCGACCGTCTCTCGGCGGAC 498  
QY 1614 tatgatgcattgccagggtccctcagtgaggacagacctgtctcttattacacagac 1673

DB 499 CGCACCACTCCACACATCGCACCACTCAACACAAACCTTTACCAACCACTCC 558  
QY 1674 gagaccggtgtgtcttctacgaccatgtccctcccccctctcgcgtccgcgggacacctg 1733  
DB 559 AACACCTCTACCCACCAACCCGCCACACACACACCCCTTCATCAGCATGTCTACCGCCACACC 618  
QY 1734 acctacactgagactgcgagaaacggtcagaataatatactcttcaactc 1781  
DB 619 ACCACCCCGACACCACTGGACCCACACATCACCACCGACCCGTC 666  
RESULT 5  
US-08-928-361B-2  
; Sequence 2, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,361B  
; FILING DATE: 12-SEP-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,062  
; FILING DATE: 13-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Verny, Hana  
; REGISTRATION NUMBER: 30,518  
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-1677  
; TELEFAX: 650-324-1678  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5511 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-928-361B-2

Query Match 2.0%; Score 45.4; DB 3; Length 5511;  
Best Local Similarity 45.2%; Pred. No. 0.0024;  
Matches 212; Conservative 0; Mismatches 251; Indels 6; Gaps 1;  
QY 1301 caagaatcacatgaactctctgaactatggtatctctatagcaagtcctcttcgagggccg 1360  
DB 926 CAACACACACACACAACTACTACTACTACTAGACACAAACAAACACACACAA 985  
QY 1361 caacgagaagggtgtctctggtcactcatgagccacacgaggaattcgaaacgga 1420  
DB 986 CAACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1045  
QY 1421 ttctctctctctac 1480  
DB 1046 CAACACACACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1105



```

;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 212-867-0123
;
; TELEFAX: 212-878-9655
;
; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
;
;     LENGTH: 1425 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;
;     MOLECULE TYPE: cDNA
;
;     HYPOTHETICAL: NO
;
;     ANTI-SENSE: NO
;
;     ORIGINAL SOURCE:
;
;     ORGANISM: Aspergillus aculeatus
;
; US-08-078-551-2

```

```
Query Match      1.9%; Score 45.2; DB 2; Length 1425;
Best Local Similarity 45.3%; Pred. No. 0.0012;
Matches 164; Conservative 0; Mismatches 198; Indels 0
```

[illegible]

## RESULT 8

```

US-09-190-982-2
: Sequence 2, Application US/09190982
: Patent No. 5998190
: GENERAL INFORMATION:
: APPLICANT: Dalboge, Henrik
: APPLICANT: Christgau, Stephan
: APPLICANT: Andersen, Lene N.
: APPLICANT: Kofod, Lene V.
: APPLICANT: Kauppinen, Sakari M.
: APPLICANT: Nielsen, Jack B.
: APPLICANT: Dambmann, Claus
: TITLE OF INVENTION: An Enzyme with Protease Activity
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5998190o No. 5998190disk of No. 5998190th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/190,982
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/378,551
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 95/02044
FILING DATE: 19-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4006,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Aspergillus aculeatus
US-09-190-982-2

```

Query Match 1.9%; Score 45.2; DB 2; Length 1425;  
Best Local Similarity 45.3%; Pred. No. 0.0012;  
Matches 164; Conservative 0; Mismatches 198; Indels 0

QY	1253	caaccocaaacgcgttgctctgagcaggaaacgcgtgctggcattgccaagaatgacga	1312
Db	481	CAGGTGAGCATCTCTACGGGACGGCAGCTCGCCAGCGCGACGCTCTACAGGACAA	540
QY	1313	tgacttcctgaaactatggttatctcagaagtcacattctcagcgcgcacacgaagagg	1372
Db	541	GGTCACCGTCGGCAGGTGACGCCACGACCCAGGCCGCTCGAGCGCCGACCGCATCAG	600
QY	1373	cggtcctggtcactaacgatggcacaacgcagagaattcgaaccggattctctcttctt	1432
Db	601	CTCGAGTTTCGTCCAGACACACGACCGACGGCTGTTGGGGTCTGGCCTTCAGCTCGAT	660
QY	1433	caactacgttaaccagacctcccggtccaacgtggcgcttgaacacttcttccaaga	1492
Db	661	CAACACGGTCTCCCGCCGGCCGACACCACCTCTTTCGACACCGTCAAGTCAGCGCTGA	720
QY	1493	cgctcactcgggtgctccctcaaatctcttaacattaccctccctcgtcgggcaccaa	1552
Db	721	CAGCCCCCTCTTCGCCGTGCAGCTCAGTACCAGCCGCCGGGTACCTTACGATTTCCGGGTT	780
QY	1553	cacaactccatgcaccccacoggtaactctctactcagtgtaggtcttcgtccaacgagat	1612
Db	781	CATCGACTCCTCCCAAGTACACCGGCTCCCTGACTTACGCCAACGTGACGAGTCCCGAGG	840
QY	1613	ct 1614	
Db	841	CT 842	

## RESULT

US-08-700-651-1 ; Sequence 1, Application US/08700651B  
; Patent No. 6015882  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN





TELEPHONE: 212-986-4090  
TELEFAX: 212-818-9479  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28804 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FRAGMENT TYPE: N-terminal  
US-08-592-874-1

Query Match 1.8%; Score 42.8; DB 2; Length 28804;  
Best Local Similarity 50.5%; Pred. No. 0.037;  
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
Qy 1487 ccaagacgctaccctggtgctccctaaattcttaccataaacccctcctgctgctg 1546  
Db 23471 CGAACAGCCTTCGCCAACGGCTTCGGGAAGCCTTCATCATCGCGCGCGATTTCGTCGG 23530  
Qy 1547 caccacacacactccatgcacccaccggttaacgtctctctacggtgaggtcttgcgtcaa 1606  
Db 23531 CAACGATCCAGCGCGCTGATCCTCGGCGACAACTATCAGCGTGAAGAGATGGCGGA 23590  
Qy 1607 gcagatctatgatgccattcgccagggccctcagtgaggagacccctgctcttcattac 1666  
Db 23591 GCCTGCCAGGCGCTCGCGCGCGCGCATCGCAGGCGCGCGGAACGTTTCGCCCTATCA 23650  
Qy 1667 ctacgacgagaccggtggtcttctacg 1692  
Db 23651 TGTCCGACGATCCCGAGCGCTACCGCG 23676

## RESULT 14

US-09-096-942-2  
Sequence 2, Application US/09096942  
Patent No. 6027925  
GENERAL INFORMATION:  
APPLICANT: Pollock, Thomas J  
APPLICANT: Mikolajczak, Marcia  
APPLICANT: Yamazaki, Motohide  
APPLICANT: Thorne, Linda  
APPLICANT: Amentrout, Richard W  
TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria  
FILE REFERENCE: seq list for appl filed from pro. appl  
CURRENT APPLICATION NUMBER: US/09/096,942  
EARLIER FILING DATE: 1998-06-12  
EARLIER FILING DATE: 1997-06-12  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 28804  
TYPE: DNA  
ORGANISM: Sphingomonas sp. S88  
US-09-096-942-2

Query Match 1.8%; Score 42.8; DB 3; Length 28804;  
Best Local Similarity 50.5%; Pred. No. 0.037;  
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
Qy 1487 ccaagacgctaccctggtgctccctaaattcttaccataaacccctcctgctgctg 1546  
Db 23471 cgaacagccttcgcccacggccttcggaagccttcattcatcgccgatttcgctcg 23530  
Qy 1547 caccacacacactccatgcacccaccggttaacgtctctctacggtgaggtcttgcgtcaa 1606  
Db 23531 caacgatccagcgctgattcctcgcgacacacatctatcacgggtgaaagatggcgga 23590  
Qy 1607 gcagatctatgatgccattcgccagggccctcagtgaggagacccctgctcttcattac 1666

Db 23591 gcgctgccagggcagctgcgcccagggcatcgagggcgcggaacgtgttcgctcatca 23650  
Qy 1667 ctacgacgagaccggtggtcttctacg 1692  
Db 23651 tgtcgacgaccccgagcgctacggcg 23676

## RESULT 15

US-09-096-867-2  
Sequence 2, Application US/09096867  
Patent No. 6030817  
GENERAL INFORMATION:  
APPLICANT: Pollock, Thomas J  
APPLICANT: Mikolajczak, Marcia  
APPLICANT: Yamazaki, Motohide  
APPLICANT: Thorne, Linda  
APPLICANT: Amentrout, Richard W  
TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria  
FILE REFERENCE: seq list for appl filed from pro. appl  
CURRENT APPLICATION NUMBER: US/09/096,867  
EARLIER FILING DATE: 1998-06-11  
EARLIER FILING DATE: 1997-06-12  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 28804  
TYPE: DNA  
ORGANISM: Sphingomonas sp. S88  
US-09-096-867-2

Query Match 1.8%; Score 42.8; DB 3; Length 28804;  
Best Local Similarity 50.5%; Pred. No. 0.037;  
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
Qy 1487 ccaagacgctaccctggtgctccctaaattcttaccataaacccctcctgctgctg 1546  
Db 23471 cgaacagccttcgcccacggccttcggaagccttcattcatcgccgatttcgctcg 23530  
Qy 1547 caccacacacactccatgcacccaccggttaacgtctctcctcagtgaggtcttgcgtcaa 1606  
Db 23531 caacgatccagcgctgattcctcgcgacacacatctatcacgggtgaaagatggcgga 23590  
Qy 1607 gcagatctatgatgccattcgccagggccctcagtgaggagacccctgctcttcattac 1666  
Db 23591 gcgctgccagggcagctgcgcccagggcatcgagggcgcggaacgtgttcgctcatca 23650  
Qy 1667 ctacgacgagaccggtggtcttctacg 1692  
Db 23651 tgtcgacgaccccgagcgctacggcg 23676

Search completed: February 26, 2001, 16:21:07  
Job time: 48340 sec

Tue Feb 27 08:38:58 2001

us-09-426-072-1.rni

Page 10

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2001, 18:58:15 ; Search time 64.74 Seconds  
(without alignments)  
129.402 Million cell updates/sec

Title: US-09-426-072-2\_COPY\_20\_264

Perfect score: 1323  
Sequence: 1 SPVTSEYTSVREAPFGYKPG.....NYDTGNGEFPDLSLFTYVN 245

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_36:\*

- 1: /cgn2\_2/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /cgn2\_2/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /cgn2\_2/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /cgn2\_2/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /cgn2\_2/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /cgn2\_2/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /cgn2\_2/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /cgn2\_2/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /cgn2\_2/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /cgn2\_2/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /cgn2\_2/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /cgn2\_2/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /cgn2\_2/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /cgn2\_2/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /cgn2\_2/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /cgn2\_2/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /cgn2\_2/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /cgn2\_2/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /cgn2\_2/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /cgn2\_2/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	92.5	7.0	556	20 Y25010	Human GPC4 protein
2	92.5	7.0	556	20 Y25915	Human GPC4 protein
3	92.5	7.0	1288	18 W55347	H. pylori ORF 14ee
4	92.5	7.0	1288	18 W55685	H. pylori ORF 07ee
5	92.5	7.0	1844	18 W18302	Phototransducin lumin
6	92.5	7.0	1844	19 W56558	Toxin TcbAii, enco
7	92.5	7.0	2504	18 W17871	Phototransducin lumin
8	92.5	7.0	2504	19 W56557	Toxin TcbA, encode
9	90.5	6.8	1041	20 Y16788	S. cerevisiae acid
10	89.5	6.8	526	13 R26174	Protein disulphide
11	88.5	6.7	556	21 Y66667	Membrane-bound pro
12	88	6.7	641	20 Y36823	Amino acid sequenc

13	86.5	6.5	772	16 R70690	Mesquite allergen1
14	86.5	6.5	1041	18 W11866	Arhi gene product,
15	86.5	6.5	1041	20 Y29830	Saccharomyces cere
16	84	6.3	1290	18 W07609	Rat von Ebner's gl
17	83.5	6.3	2893	19 W98828	H. pylori GPO 148
18	83.5	6.3	2893	19 W71556	Helicobacter poly
19	83	6.3	940	11 R70700	Fibronectin-blidin
20	83	6.3	1026	21 Y83025	Staufen protein of
21	82.5	6.2	521	12 R13610	Ala(147), Ala(189)
22	82.5	6.2	521	12 R14105	Ala(189) neutral p
23	82.5	6.2	522	20 W78488	Human R1P140 ligan
24	82.5	6.2	594	20 Y17899	Methanococcus jann
25	82.5	6.2	594	21 Y52017	M. jannaschii MJ07
26	82.5	6.2	594	21 Y51646	M. jannaschii MJ07
27	82.5	6.2	1721	19 W48299	Cryptosporidium pa
28	82	6.2	323	20 Y37770	Chlamydia trachoma
29	82	6.2	559	18 W89803	Staphylococcus aur
30	82	6.2	1096	17 W01596	Chlamydia trachoma
31	81.5	6.2	223	19 W57422	Inositol-1,4,5-tri
32	81	6.1	746	17 R93270	Amino acid sequenc
33	80.5	6.1	257	20 Y19890	B. burgdorferi ant
34	80.5	6.1	257	20 Y19891	B. burgdorferi ant
35	80.5	6.1	428	12 R10427	Rice chloroplast
36	80.5	6.1	587	17 R88656	N. meningitidis 828
37	80	6.0	372	19 W47095	Murine DMP1 DNA bl
38	80	6.0	556	20 Y13453	Amino acid sequenc
39	80	6.0	761	19 W47093	Murine cyclin D tr
40	79.5	6.0	300	18 W27278	Bacillus subtilis
41	79.5	6.0	521	7 P61426	Sequence of neutra
42	79.5	6.0	521	8 P71061	Neutral protease f
43	79.5	6.0	521	8 P70054	Bacillus subtilis
44	79.5	6.0	521	11 R03738	Carboxyl hydrolase
45	79.5	6.0	521	12 R14104	Ala(147) neutral p

## ALIGNMENTS

RESULT 1  
Y25910  
ID Y25910 standard; Protein; 556 AA.  
XX AC Y25910;

DT 06-OCT-1999 (first entry)

XX Human GPC4 protein.

XX Glypican; GPC1; GPC3; GPC4; GPC5; GPC6; human; glypican-related protein;  
KW glypican-6; glypican-4; glypican-1; glypican-3; glypican-5; diagnosis;  
KW treatment; abnormal; cell growth; cell behaviour; somatic overgrowth;  
KW tumour formation.

XX Homo sapiens.

XX WO9937764-A2.

XX 29-JUL-1999.

XX 20-JAN-1999; 99WO-EP00329.

XX 27-JAN-1998; 98EP-0200226.

XX (VLAAS) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

XX David GJF, Veugelers MPD;

XX WPI: 1999-469128/39.

XX N-PSDB; 200684.

XX New polynucleotides encoding glypican-related proteins, used to  
diagnose, e.g. tumor formation

PS Example 2; Fig 2; 79pp; English.  
XX This invention describes the isolation of novel human polynucleotides  
CC encoding glypican-related proteins, glypican-6 (GPC6) and glypican-4  
CC (GPC4). The invention also describes the polynucleotide and encoded  
CC protein sequences of glypican-1 (GPC1), glypican-3 (GPC3) and glypican-5  
CC (GPC5). The products of the invention can be used to diagnose and treat  
CC disorders and diseases, particularly those involving abnormal cell  
CC growth and behaviour, such as somatic overgrowth and tumour formation.  
CC This sequence represents the human GPC4 protein.  
XX  
SQ Sequence 556 AA;

Query Match 7.0%; Score 92.5; DB 20; Length 556;  
Best Local Similarity 25.8%; Pred. No. 0.63;  
Matches 54; Conservative 29; Mismatches 83; Indels 43; Gaps 13;  
QY 47 NILGG-VRROG-LDNPINN-----GPFCKYKASDPSSGKYCTQAKDYDSVF 91  
DB 272 nimgclangqgldfewnfdamlnvaerlegpf-niesvmdp-----idvklisdaim 324  
QY 92 NDPDHSVTGNLEFYGTTPNGAIASGV--VADQSGFLNAQLNDYPKLAPEEATROVM 149  
DB 325 nmqdnsvqskvfgcgppk--plpagrirsisesaf-sarfrph---hpeertptaa 378  
QY 150 GYYTEERVPTLVDLVEFTTFNSWFCVPGTPNRLCALAGTAAGKNDODFLNYGIS 209  
DB 379 gtsidrlvtvdvdklkqakf--wssl-----psnvcndermaagng-neddcwngkgk 429  
QY 210 SKSIFEAAANEKGVSWLNYDGTNGEFEPDS 238  
DB 430 srylfavtnglgnv----qgnnpevqvdt 454

RESULT 2  
Y25915  
ID Y25915 standard; Protein; 556 AA.  
AC Y25915;  
DT 06-OCT-1999 (first entry)  
DE Human GPC4 protein fragment.  
KW Glypican; GPC1; GPC3; GPC4; GPC5; GPC6; human; glypican-related protein;  
KW glypican-6; glypican-4; glypican-1; glypican-3; glypican-5; diagnosis;  
KW treatment; abnormal; cell growth; cell behaviour; somatic overgrowth;  
KW tumour formation.  
XX Homo sapiens.  
XX WO9937764-A2.  
XX 29-JUL-1999.  
XX 20-JAN-1999; 99WO-EP00329.  
XX 27-JAN-1998; 98EP-0200226.  
XX (VLAAR-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.  
XX David GJF, Veugelers MPD;  
XX WPI: 1999-469128/39.  
XX New polynucleotides encoding glypican-related proteins, used to  
XX diagnose, e.g. tumor formation  
XX  
XX Disclosure; Fig 6; 79pp; English.  
XX This invention describes the isolation of novel human polynucleotides  
CC encoding glypican-related proteins, glypican-6 (GPC6) and glypican-4

CC (GPC4). The invention also describes the polynucleotide and encoded  
CC protein sequences of glypican-1 (GPC1), glypican-3 (GPC3) and glypican-5  
CC (GPC5). The products of the invention can be used to diagnose and treat  
CC disorders and diseases, particularly those involving abnormal cell  
CC growth and behaviour, such as somatic overgrowth and tumour formation.  
CC This sequence represents a human GPC4 protein fragment.  
XX  
SQ Sequence 556 AA;

Query Match 7.0%; Score 92.5; DB 20; Length 556;  
Best Local Similarity 25.8%; Pred. No. 0.63;  
Matches 54; Conservative 29; Mismatches 83; Indels 43; Gaps 13;  
QY 47 NILGG-VRROG-LDNPINN-----GPFCKYKASDPSSGKYCTQAKDYDSVF 91  
DB 272 nimgclangqgldfewnfdamlnvaerlegpf-niesvmdp-----idvklisdaim 324  
QY 92 NDPDHSVTGNLEFYGTTPNGAIASGV--VADQSGFLNAQLNDYPKLAPEEATROVM 149  
DB 325 nmqdnsvqskvfgcgppk--plpagrirsisesaf-sarfrph---hpeertptaa 378  
QY 150 GYYTEERVPTLVDLVEFTTFNSWFCVPGTPNRLCALAGTAAGKNDODFLNYGIS 209  
DB 379 gtsidrlvtvdvdklkqakf--wssl-----psnvcndermaagng-neddcwngkgk 429  
QY 210 SKSIFEAAANEKGVSWLNYDGTNGEFEPDS 238  
DB 430 srylfavtnglgnv----qgnnpevqvdt 454

RESULT 3  
W55547  
ID W55547 standard; Protein; 1288 AA.  
XX  
AC W55547;  
DT 24-JUN-1998 (first entry)  
DE H. pylori ORF 14ee41924\_2458267\_c2\_93 secreted protein.  
XX  
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
XX  
XX Helicobacter pylori.  
XX WO9737044-A1.  
XX  
XX 09-OCT-1997.  
XX  
XX 27-MAR-1997; 97WO-US05223.  
XX  
XX 06-DEC-1996; 96US-0761318.  
XX 29-MAR-1996; 96US-0625811.  
XX 02-APR-1996; 96US-0758731.  
XX 25-OCT-1996; 96US-0736905.  
XX 28-OCT-1996; 96US-0738859.  
XX  
XX (ASTR ) ASTRA AB.  
XX  
XX Alm RA, Smith D;  
XX WPI: 1997-503122/46.  
XX N-PSDB; V24956.  
XX  
XX Helicobacter pylori nucleic acid sequences and encoded  
XX polypeptide(s) useful in vaccines to treat or prevent H. pylori  
XX infection and for diagnosis of H. pylori infection  
XX  
XX Disclosure; Page 750-753; 1145pp; English.  
XX  
XX This sequence is a H. pylori secreted protein. The protein may be used

PR 02-APR-1996; 96US-0758731.

XX  
DT 30-JAN-1998 (first entry)



CC protein. The native toxins are secreted proteins. The proteins are  
 CC toxic to insects upon exposure and especially when ingested. The  
 CC nucleic acid sequence can be used to produce transgenic plants,  
 CC baculoviruses or microbial hosts for toxin production. They can be used  
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
 CC codling moth, corn earworm, European corn borer or tobacco hornworm  
 XX  
 SQ Sequence 1844 AA;

Query Match 7.0%; Score 92.5; DB 19; Length 1844;  
 Best Local Similarity 24.1%; Pred. No. 4;  
 Matches 46; Conservative 30; Mismatches 62; Indels 53; Gaps 11;  
 QY 35 IWLLILENRSFDN-----ILGVRROGLDNPINNGPFC-----NYKNASD-- 74  
 Db 1067 liwleqsksdgkttiygynklahirydgsnt----pffdtvtekvknytsstdaa 1122  
 QY 75 PSSGKYCTQAKDYDS---VFNDPDHSHVTGNLFYGTYPNNGAIAAGKVVADQS--GFL 129  
 Db 1123 eslglyctgyqgdettllvmfysmqss-----yssytdnnapvlglyifadmsdmt 1174  
 QY 130 NAQL-----NDYPK----LAPEEA-----TRQVMGYTTE-EVPTLVLDVDEFTTNSW 173  
 Db 1175 naqatynwnsyppqfdtvmadpdsdnkvtirvnnryaedyelpsvtsnsnywgghs 1234  
 QY 174 FSCVPGPTNPN 184  
 Db 1235 ltmylgsgvyn 1245

## RESULT 7

ID W17871  
 XX W17871 standard; Protein; 2504 AA.  
 AC W17871;  
 XX  
 DT 29-JAN-1998 (first entry)  
 XX  
 DE Photorhabdus luminescens insect toxin protein TcBA.  
 XX  
 KW Insecticide; insect; toxin; pest control; biological control;  
 KW Photorhabdus luminescens; TcBA; Southern corn rootworm;  
 KW Colorado potato beetle; Western corn rootworm; meal worm;  
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
 KW Diptera, Dictyoptera; Acarina; Homoptera.  
 XX  
 OS Photorhabdus luminescens strain W-14 (ATCC 55397).  
 XX  
 PN W09717432-Al.  
 XX  
 PD 15-MAY-1997.  
 XX  
 PF 06-NOV-1996; 96WO-US18003.  
 XX  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1995; 95US-0007255.  
 PR 28-FEB-1996; 96US-0608423.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA.  
 XX  
 PF WPI: 1997-281022/25.

DR N-PSDB; T68836.

XX  
 PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
 PT can be genetically engineered into insect larvae food and plants for  
 PT insect control

XX Claim 8; Page 119-129; 276pp; English.

XX  
 CC This polypeptide comprises a specifically claimed 280.6 kDa  
 CC insecticidal toxin, TcBA, of Photorhabdus luminescens that is  
 CC proteolytically cleaved into 207.6 kDa (see W18302) and 62.9 kDa  
 CC polypeptides (see W18303). The TcBA polypeptide can be expressed  
 CC in host cells using a gene (see T68836) isolated from a genomic  
 CC library. Claimed toxin proteins of P. luminescens (see W1871,  
 CC W17884-89, W17899-900, W18301-06) can be applied to, or genetically  
 CC engineered into, insect larvae food and plants for insect control.  
 CC The Photorhabdus toxins are particularly effective against Southern  
 CC corn rootworm, Colorado potato beetle, Western corn rootworm, meal  
 CC worm, boll weevil and turf grub (Coleoptera), beet armyworm, black  
 CC cutworm, cabbage looper, codling moth, corn earworm, European corn  
 CC borer, tobacco hornworm and tobacco budworm (Lepidoptera), and  
 CC are also active against insects of the orders Hymenoptera, Diptera  
 CC Dictyoptera, Acarina and Homoptera. (All claimed).  
 XX  
 SQ Sequence 2504 AA;

Query Match 7.0%; Score 92.5; DB 18; Length 2504;  
 Best Local Similarity 24.1%; Pred. No. 6.4;  
 Matches 46; Conservative 30; Mismatches 62; Indels 53; Gaps 11;

QY 35 IWLLILENRSFDN-----ILGVRROGLDNPINNGPFC-----NYKNASD-- 74  
 Db 1154 liwleqsksdgkttiygynklahirydgsnt----pffdtvtekvknytsstdaa 1209  
 QY 75 PSSGKYCTQAKDYDS---VFNDPDHSHVTGNLFYGTYPNNGAIAAGKVVADQS--GFL 129  
 Db 1210 eslglyctgyqgdettllvmfysmqss-----yssytdnnapvlglyifadmsdmt 1261  
 QY 130 NAQL-----NDYPK----LAPEEA-----TRQVMGYTTE-EVPTLVLDVDEFTTNSW 173  
 Db 1262 naqatynwnsyppqfdtvmadpdsdnkvtirvnnryaedyelpsvtsnsnywgghs 1321  
 QY 174 FSCVPGPTNPN 184  
 Db 1322 ltmylgsgvyn 1332

## RESULT 8

W56557  
 ID W56557 standard; Protein; 2504 AA.  
 XX  
 AC W56557;  
 XX  
 DT 07-AUG-1998 (first entry)  
 XX  
 DE Toxin TcBA, encoded by the tcbA gene from genomic region tcb.  
 XX  
 KW Photorhabdus luminescens W-14; nematode; symbiotic;  
 KW Heterorhabditis; tcb; tcb; tcc; tcd; insecticidal activity; toxin;  
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW Tobacco hornworm; budworm.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN W09808932-Al.  
 XX  
 PD 05-MAR-1998.  
 XX  
 PF 05-MAY-1997; 97WO-US07657.

XX 06-NOV-1996; 96W0-US18003.  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1996; 96US-0743699.  
 XX  
 PA (DWSC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DI, Orr GL;  
 PI Petrelli J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA, Sukhapinda K;  
 XX  
 DR WPI; 1998-179427/16.  
 DR N-PSDB; V29985.  
 XX  
 XX Isolated toxins from Photorhabdus luminescens strains - useful for  
 PT control of insect pests  
 XX  
 XX Claim 34; Pages 163-169; 321pp; English.  
 XX  
 CC The present sequence represents a protein named TcbA of the bacterium  
 CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
 CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
 CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
 CC produced from these regions that are associated with insecticidal  
 CC activity. The native toxins are secreted proteins. The proteins are  
 CC toxic to insects upon exposure and especially when ingested. The  
 CC nucleic acid sequence can be used to produce transgenic plants,  
 CC baculoviruses or microbial hosts for toxin production. They can be used  
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
 CC codling moth, corn earworm, European corn borer or tobacco hornworm  
 CC or budworm.  
 XX  
 SQ Sequence 2504 AA;

Query Match 7.0%; Score 92.5; DB 19; Length 2504;  
 Best Local Similarity 24.1%; Pred. No. 6.4;  
 Matches 46; Conservative 30; Mismatches 62; Indels 53; Gaps 11;

QY 35 IYWLILENRSFN-----ILGGVVRQGLDNPINNGPFC-----NYKNASD-- 74  
 Db 1154 ILWEGSKKSDGKTTIYQYKLNKLIHRYDGSWLT-----PTFDVTEKVKNYTSIDAA 1209  
 QY 75 PSSGKYCTOAKDYDS---VFNDPDHSHVTGNLIFYGTYPNNGATASGVVADQS--GFL 129  
 Db 1210 eslglyctgyggedtllvmfysmqss-----ysaytdnnapvtglyifadmsdnt 1261  
 QY 130 NAOL-----NDYPK---LAPEEA-----TROVMGYTTEE-EVPTLVLDVDEFTFNSW 173  
 Db 1262 nagatnynwnsyqfdvtvmdpdsdnkktvtrvnnryadyeipssvtsnysnyswgds 1321  
 QY 174 FSCVPGPTNP 184  
 Db 1322 ltmlygsgvnp 1332

RESULT 9  
 Y16788  
 ID Y16788 standard; Protein; 1041 AA.  
 XX  
 AC Y16788;  
 XX  
 DT 27-JUL-1999 (first entry)  
 XX  
 DE S. cerevisiae acidic trehalase (ATH1) gene product.  
 XX  
 XX Acidic trehalase; ATH1 gene; gene manipulation; baker's yeast; dough;  
 KW fermentation; baking; freezing; bread; trehalose degradation.

XX Saccharomyces cerevisiae.  
 XX  
 PN EP921190-A2.  
 XX  
 PD 09-JUN-1999.  
 XX  
 PF 04-DEC-1998; 98EP-0403046.  
 XX  
 PR 08-DEC-1997; 97JP-0352016.  
 XX  
 PA (NORO ) NAT FOOD RES INST MIN AGRIC.  
 PA (ORIY ) ORIENTAL YEAST CO LTD.  
 XX  
 PI Iyo C, Mori K, Nakajima R, Shima J, Suzuki Y, Takano H;  
 PI Watanabe H;  
 XX  
 DR WPI; 1999-315405/27.  
 DR N-PSDB; X60600.  
 XX  
 PT New acid trehalase (ATH1) gene-disrupted haploid yeast, useful for  
 PT preparing high-sugar dough resistant to freezing  
 XX  
 PS Examples; Page 16-23; 57pp; English.  
 XX  
 CC The invention relates to an acid trehalase (ATH1) gene-disrupted haploid  
 CC yeast produced by gene manipulation, where the diploid is a practical  
 CC baker's yeast. The new yeast strain is useful in dough which, upon  
 CC subsequent fermentation and baking after a period of freezing, produces  
 CC a high quality bread or sweetened bun. The new yeasts have a reduced  
 CC ability of trehalase degradation during fermentation and thus an  
 CC increased intracellular osmotic pressure, providing resistance to long  
 CC term freeze-storage when used in low and high-sugar dough. The dough  
 CC makes better bread than dough using prior art baker's yeast. The new  
 CC strain and ATH1 non-disrupted gene strains were used to prepare dough,  
 CC and the new strain showed resistance to freezing. The present sequence  
 CC represents an ATH1 gene product.  
 XX  
 SQ Sequence 1041 AA;

Query Match 6.8%; Score 90.5; DB 20; Length 1041;  
 Best Local Similarity 20.8%; Pred. No. 2.7;  
 Matches 60; Conservative 46; Mismatches 105; Indels 77; Gaps 15;

QY 3 VFSEY---TSVREAPFGYKPGSKSINLKDKNVENIWLLENRS-----FONILGG 51  
 Db 446 vdveyhinvdvasfsiylnghegid--eylryttplknaaqftaykynssigl 503  
 QY 52 VRRQGLNP-----INNGPFCNKNASDPSSGKYCTQAKDYDSVFNDPDHSHVTGNLEF 105  
 Db 504 yetynltldpdefanhnngaf---tnagiktlkwtadighlgevdpkwsiskdl-- 558  
 QY 106 YGTYPNNGATASGVVADQSGFLNAOLNDYKPLAPEEATROV--MGYYTTEEVPTLVDL 163  
 Db 559 ---ylypr-----ssnltleysg-----mmsveikqadvltlmvplyindesi--lnna 604  
 QY 164 VDEFTFNSWFCV-EGPTNPNRLCALAGTAAGHG-----K 198  
 Db 605 ikdlyyyserqsasgpmtypvfvaagaag-llnhgsssqsylyksvlylrapfaqseq 663  
 QY 199 NDDDFLNYGISSKSI-----FEAANEKGVSWLNYDGTNGEPEPS 238  
 Db 664 sadnflngltqpaflltanggfllqsnlglgiry---syevdpdt 708

RESULT 10  
 R26174  
 ID R26174 standard; Protein; 526 AA.  
 XX  
 AC R26174;  
 XX  
 DT 04-FEB-1993 (first entry)

```

xx DE Protein disulphide isomerase.
xx KW Yeast; isomerisation; catalyst; enzyme; disulphide bonds.
xx OS Saccharomyces cerevisiae.
xx PN JP04197176-A.
xx PD 16-JUL-1992.
xx PF 28-NOV-1990; 90JP-0322620.
xx PR 28-NOV-1990; 90JP-0322620.
xx PA (TOFU) TONEN CORP.
xx WI; 1992-288434/35.
xx DR N-PSDB; Q27388.
xx PT Yeast protein di-sulphide isomerase - used to give active
xx PT conformation to protein having inadequate di-sulphide bond
xx PS Disclosure; Fig 15; 22pp; Japanese.
xx CC The sequence given is a yeast protein disulphide isomerase (PDI)
xx CC which catalyses the isomerisation of disulphide bonds. It has a
xx CC molecular weight of approx 70 kD and has an optimum pH for activity
xx CC of 8.75. This enzyme can be used to give active conformation to a
xx CC protein having inadequate disulphide bonds. The gene encoding this
xx CC protein can be used to produce expression vectors for the large scale
xx CC production of PDI.
xx SQ Sequence 526 AA;

Query Match
Best Local Similarity 22.6%; Pred. No. 1.2;
Matches 56; Conservative 39; Mismatches 94; Indels 59; Gaps 13;

QY 2 PVTSEYTSREAPFGYKPGSKESIEINKDKVENIWMLEINRSFDNIGVRRQGLDNPI 61
DQ : : : : : l : l : l : l : : : : : : : : : : : : : : : :
Db 312 plfaidmtdedkyglpqlseafdelssdkl-----vleskaieslvkdfik-gdaspi 364
QY 62 NNGPFCNYKNASDSSGKYCTQAKDYDSVFNPDHSVTGNLFYCTYTPNNGATASGV 121
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 vk-----sqeifengqssvflqgknhdvndpkkdv-----lvly--yap----- 404
QY 122 VADSGFLNAQLNDYPKLAP--EEATROVMGY---YTBEEVPTLVDLDFEFTFNSWFSC 176
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 405 ---wcghckrlaptyqeladyanatsvliakldhtendvr--gvvlevtqgssytqv 458
QY 177 VPGFTPNRLCALAGTAGHGKNDPFLNYGTS-----KSIFEAAEKGVSWLNVDGT 230
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 v---rspnllct-----kvqdpwtlytsrkrtdvdkalyeaqekaeeead-- 504
QY 231 NGEPFEDPS 238
DQ : : : : :
Db 505 -aaeada 511

RESULT 11
Y66667
ID Y66667 standard; protein; 556 AA.
XX AC
XX Y66667;
XX DT
XX 05-APR-2000 (first entry)
XX DE Membrane-bound protein PRO937.
KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping.

```

```

XX OS Homo sapiens.
XX PN W09963088-A2.
XX PD 09-DEC-1999.
XX PF 02-JUN-1999;
XX PR 02-JUN-1998; 99WO-US12252.
XX PR 02-JUN-1998; 98US-0087607.
XX PR 02-JUN-1998; 98US-0087609.
XX PR 02-JUN-1998; 98US-0087759.
XX PR 03-JUN-1998; 98US-0087827.
XX PR 04-JUN-1998; 98US-0088021.
XX PR 04-JUN-1998; 98US-0088025.
XX PR 04-JUN-1998; 98US-0088028.
XX PR 04-JUN-1998; 98US-0088029.
XX PR 04-JUN-1998; 98US-0088030.
XX PR 04-JUN-1998; 98US-0088033.
XX PR 04-JUN-1998; 98US-0088326.
XX PR 05-JUN-1998; 98US-0088167.
XX PR 05-JUN-1998; 98US-0088202.
XX PR 05-JUN-1998; 98US-0088212.
XX PR 05-JUN-1998; 98US-0088217.
XX PR 09-JUN-1998; 98US-0088655.
XX PR 10-JUN-1998; 98US-0088722.
XX PR 10-JUN-1998; 98US-0088730.
XX PR 10-JUN-1998; 98US-0088734.
XX PR 10-JUN-1998; 98US-0088738.
XX PR 10-JUN-1998; 98US-0088740.
XX PR 10-JUN-1998; 98US-0088741.
XX PR 10-JUN-1998; 98US-0088742.
XX PR 10-JUN-1998; 98US-0088810.
XX PR 10-JUN-1998; 98US-0088811.
XX PR 10-JUN-1998; 98US-0088824.
XX PR 10-JUN-1998; 98US-0088825.
XX PR 11-JUN-1998; 98US-0088826.
XX PR 11-JUN-1998; 98US-0088858.
XX PR 11-JUN-1998; 98US-0088861.
XX PR 11-JUN-1998; 98US-0088863.
XX PR 11-JUN-1998; 98US-0088876.
XX PR 12-JUN-1998; 98US-0089090.
XX PR 12-JUN-1998; 98US-0089105.
XX PR 16-JUN-1998; 98US-0089440.
XX PR 16-JUN-1998; 98US-0089512.
XX PR 16-JUN-1998; 98US-0089514.
XX PR 17-JUN-1998; 98US-0089532.
XX PR 17-JUN-1998; 98US-0089538.
XX PR 17-JUN-1998; 98US-0089598.
XX PR 17-JUN-1998; 98US-0089599.
XX PR 17-JUN-1998; 98US-0089600.
XX PR 18-JUN-1998; 98US-0089653.
XX PR 18-JUN-1998; 98US-0089801.
XX PR 18-JUN-1998; 98US-0089907.
XX PR 19-JUN-1998; 98US-0089908.
XX PR 19-JUN-1998; 98US-0089947.
XX PR 19-JUN-1998; 98US-0089948.
XX PR 19-JUN-1998; 98US-0089952.
XX PR 22-JUN-1998; 98US-0090246.
XX PR 22-JUN-1998; 98US-0090252.
XX PR 22-JUN-1998; 98US-0090254.
XX PR 23-JUN-1998; 98US-0090349.
XX PR 23-JUN-1998; 98US-0090355.
XX PR 24-JUN-1998; 98US-0090429.
XX PR 24-JUN-1998; 98US-0090431.
XX PR 24-JUN-1998; 98US-0090435.
XX PR 24-JUN-1998; 98US-0090444.
XX PR 24-JUN-1998; 98US-0090445.
XX PR 24-JUN-1998; 98US-0090461.
XX PR 24-JUN-1998; 98US-0090472.
XX PR 24-JUN-1998; 98US-0090535.
XX PR 24-JUN-1998; 98US-0090538.
XX PR 24-JUN-1998; 98US-0090540.

```





XX 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97ER-0015041.  
 PR 17-DEC-1997; 97ER-0016034.  
 XX (GEST ) GENSET.  
 XX Griffals R;  
 PI WPI: 1999-371125/31.  
 XX Genome sequence of Chlamydia trachomatis  
 XX Disclosure: Page 704; 1755pp; English.  
 XX Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of  
 CC Chlamydia trachomatis (see Z01425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases. Chlamydia  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perihepatitis, Bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 XX  
 SQ Sequence 641 AA;  
 Query Match 6.7%; Score 88; DB 20; Length 641;  
 Best Local Similarity 26.3%; Pred. No. 2.3;  
 Matches 46; Conservative 22; Mismatches 69; Indels 38; Gaps 11;  
 QY 37 WLI--LENRSFNIL-----GCVRRQGLDNPINNGPCNY--KNASDPSSGKYCTQ-- 83  
 ||| | ||||| || : : : : : ||| : : ||| : :  
 Db 289 wllpcfekrafdnllmlyvteagvymrnpelfvivakrvlnylkeltopksgsfylsey 348  
 QY 84 -----AKDYDSVF---NDPDHVTGNLLE-FYGTTPNNGAIASGKVVADQSGFLN---- 130  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 349 gqvwagtdldsytwsgelrsllegeamfeydvsvregiongrnllhvspymnrkei 408  
 QY 131 -----AQLNDYKPLAPEATQVMGYTTEEEVPTLVLDVDEFTTF-NSW--PSCV 177  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 409 eeryhcsleefqkkl--easreklyvregktqaskd--dqsfthgngwgifslil 459  
 RESULT 13  
 R70690  
 ID R70690 standard; Protein; 772 AA.  
 XX  
 AC R70690;  
 XX  
 DT 10-AUG-1995 (first entry)  
 XX  
 DE Mesquite allergenic pollen serine protease.  
 XX  
 KW Pollen; allergen; serine protease; enzyme; vaccine; diagnosis;  
 KW protease-inhibitor; enzyme-inhibitor; antiallergic.  
 XX  
 OS Mesquite.  
 XX  
 PN WO9500850-A.  
 XX  
 PD 05-JAN-1995.  
 XX  
 PF 17-JUN-1994; 94WO-US06880.  
 XX  
 PR 18-JUN-1993; 93US-0079539.  
 XX  
 PA (LXRB-) LXR BIOTECHNOLOGY INC.  
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
 XX

PI Barr PJ, Travis J;  
 XX WPI: 1995-052229/07.  
 DR N-PSDB; Q85032.  
 XX  
 PT New proteinase(s) isolated from allergenic pollen - and nucleic  
 PT acid encoding them, useful in vaccines, for diagnosis and for  
 PT identification of therapeutic proteinase inhibitors  
 XX  
 PS Claim 6; Page 56; 75pp; English.  
 XX  
 CC This protein (88 kDa) may be expressed in a recombinant yeast  
 CC using vector plasmid MPPVY. The serine protease may be used to  
 CC monitor exposure of animals to the protease, for diagnosis or for  
 CC following treatment, to immunize animals against allergic response  
 CC to pollen, etc. There are 3 potential N-linked glycosylation sites,  
 CC a potential signal sequence of 29 AAs at the N-terminus and a 200 AA  
 CC stretch at the C-terminus with 63% identity to E. coli protease-II.  
 XX  
 SQ Sequence 772 AA;  
 Query Match 6.5%; Score 86.5; DB 16; Length 772;  
 Best Local Similarity 20.4%; Pred. No. 4.3;  
 Matches 47; Conservative 35; Mismatches 81; Indels 67; Gaps 13;  
 QY 41 ENRSFDNILGGVRRQGLDNPINNGPCNY-KNASDPSSGKYCTQ-----AKDYDSVF 91  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 128 edklfaelgrikedttcaplrkgyyyvrtlagkeyaqycrrpyddkatpslydtvp 187  
 QY 92 NDP-----DHSVTGNLLEFY--GTY--TPNNGAIASGKVVADQSG-----FLNA 131  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 188 tepdapeehwildenikqnheynigafkvspnnklvayae---dtkgdelytiyvida 244  
 QY 132 QLND-----YPLAPEATQVMGYTTEEEVPTLVLDVDEFTTFNSWFCVPGP 180  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 245 etgdpigepihnvtysiewagdea---lvyitmdel-----lrpdkawhlrl-gt 290  
 QY 181 TPNRRLCAGTAACHGKNDDDF---LNYGISSKSIFFAANEKGVSNLNY 227  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 Db 291 eqskdicl-----yvekdokfslldqaseskkylfvaseskntrfnfy 333  
 RESULT 14  
 W11866  
 ID W11866 standard; Protein; 1041 AA.  
 XX  
 AC W11866;  
 XX  
 DT 18-APR-1997 (first entry)  
 XX  
 DE Ath1 gene product, Ath1p.  
 XX  
 KW Yeast Ath1; Ath1p; vacuolar acid trehalase; hydrolysis; trehalose;  
 KW baking; frozen dough; dehydrated yeast; brewing; ethanol fuel;  
 KW molasses; corn syrup; freezing; wine; fermentation.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN US5587290-A.  
 XX  
 PD 24-DEC-1996.  
 XX  
 PF 26-JUN-1995; 95US-0494714.  
 XX  
 PR 26-JUN-1995; 95US-0494714.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Destruelle M, Holzer H, Klionsky D;  
 XX WPI: 1997-064793/06.  
 DR N-PSDB; T61372.  
 DR

XX Mutant yeast deficient in prodn. of vacuolar acid trehalase - has  
 PT improved stress tolerance, esp. for dehydration, freezing and  
 PT alcohol concn., useful in baking, fermentation, etc.  
 XX  
 XX Disclosure: Column 17-24; 17pp; English.  
 XX  
 XX This sequence is encoded by the yeast ATH1 coding sequence and represents  
 CC Ath1p which is a vacuolar acid trehalase. Yeast mutants which are  
 CC deficient in the production of functional Ath1p have reduced ability  
 CC to hydrolyse trehalose. These mutants are useful in baking (partic. in  
 CC frozen dough and dehydrated yeast products), in brewing, as a source of  
 CC trehalose (used as a protectant in foods and pharmaceuticals) and for  
 CC production of ethanol fuel from molasses or corn syrup. The yeast  
 CC mutants have improved tolerance of dehydration and freezing. They can  
 CC tolerate higher levels of ethanol and can grow at high cell densities  
 CC over a range of fermentable sugar concentrations, and so can produce  
 CC drier wines and complete fermentation more quickly.  
 XX  
 XX Sequence 1041 AA;  
 SQ

Query Match 6.5%; Score 86.5; DB 18; Length 1041;  
 Best Local Similarity 21.2%; Pred. No. 6.8;  
 Matches 61; Conservative 45; Mismatches 105; Indels 77; Gaps 16;  
 QY 3 VTSEY---TSVRAPFGYKPGSKESIKENLKDVENIWLILENRS-----FDNILGG 51  
 Db 446 vdyeyhinvdvnasfsylnghegidd--eylryttwpiiknaaqfftaykvynssigl 503  
 QY 52 VRRQGLDNP-----INNGPFCNYKNASDPSSGKYCTQAKDYDSVFNDDPHSVTGNLLEF 105  
 Db 504 yetynltdpdefanhnngaf---tnagiktlkwtadgihlgevvdvdkwseiskdi-- 558  
 QY 106 YGTYTPNNGATASGVVADQSGFLNAQLNDYPKLAPEEATROV--MGYYTEEEVPTLVDL 163  
 Db 559 ---ypr-----ssnltleysg-----mmsvveikqadvltmvpilgyindesi--lnna 604  
 QY 164 VDEFTTFNSWFCV-PGPTNPNRLCALAGTAAGHG-----K 198  
 Db 605 ikdlyyyserqsgasgpamtypfvvaaag-llnhgsssqsylyksvipylrapafqseq 663  
 QY 199 NDDDFLNYGSSKSI-FAANEK-----GVSWLNYDGTNGEPEPDS 238  
 Db 664 sddnflngltqapffltanggflqsilgtgiry---syevdpdt 708

RESULT 15  
 Y29830  
 ID Y29830 standard; Protein; 1041 AA.  
 XX  
 AC Y29830;  
 XX  
 DT 15-NOV-1999 (first entry)  
 XX  
 DE Saccharomyces cerevisiae ATH1 protein.  
 XX  
 KW Saccharomyces cerevisiae; baker's yeast; neutral trehalase; NTH1;  
 KW drying-resistant; ATH1; acidic trehalase; desiccation; bread.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN EP945511-A2.  
 XX  
 PD 29-SEP-1999.  
 XX  
 PF 24-FEB-1999; 99EP-0103560.  
 XX  
 PR 04-MAR-1998; 98JP-0067672.  
 XX  
 PA (NORQ ) NAT FOOD RES INST MIN AGRIC.  
 PA (ORII ) ORIENTAL YEAST CO LTD.  
 XX

PI Mori K, Nakajima R, Shima J, Suzuki Y, Takano H;  
 PI Watanabe H;  
 XX  
 XX WPI: 1999-530049/45.  
 DR N-PSDB: Z07051.  
 XX  
 XX New dried baker's yeast, with greater resistance to desiccation  
 PT  
 XX Disclosure: Page 34-38; 56pp; English.  
 XX  
 XX The present invention describes a diploid dry baker's yeast (I) which  
 CC is obtained by mating an a-type haploid yeast whose ATH1 (acidic  
 CC trehalase gene) or NTH1 (neutral trehalase gene) gene is disrupted with  
 CC an alpha-type haploid yeast disrupted in the same gene. Both haploid  
 CC strains are derived from practical baker's yeast. The present sequence  
 CC represents ATH1 from baker's yeast (Saccharomyces cerevisiae). Yeast  
 CC requires trehalose during rehydration after a period of desiccation.  
 CC (I) is mutated in either the ATH1 or NTH1 gene which increases  
 CC intracellular trehalose and is derived from parent baker's yeast  
 CC strains which already have resistance to drying. Therefore (I) has  
 CC greater resistance to desiccation.  
 XX  
 XX Sequence 1041 AA;  
 SQ

Query Match 6.5%; Score 86.5; DB 20; Length 1041;  
 Best Local Similarity 21.2%; Pred. No. 6.8;  
 Matches 61; Conservative 45; Mismatches 105; Indels 77; Gaps 16;  
 QY 3 VTSEY---TSVRAPFGYKPGSKESIKENLKDVENIWLILENRS-----FDNILGG 51  
 Db 446 vdyeyhinvdvnasfsylnghegidd--eylryttwpiiknaaqfftaykvynssigl 503  
 QY 52 VRRQGLDNP-----INNGPFCNYKNASDPSSGKYCTQAKDYDSVFNDDPHSVTGNLLEF 105  
 Db 504 yetynltdpdefanhnngaf---tnagiktlkwtadgihlgevvdvdkwseiskdi-- 558  
 QY 106 YGTYTPNNGATASGVVADQSGFLNAQLNDYPKLAPEEATROV--MGYYTEEEVPTLVDL 163  
 Db 559 ---ypr-----ssnltleysg-----mmsvveikqadvltmvpilgyindesi--lnna 604  
 QY 164 VDEFTTFNSWFCV-PGPTNPNRLCALAGTAAGHG-----K 198  
 Db 605 ikdlyyyserqsgasgpamtypfvvaaag-llnhgsssqsylyksvipylrapafqseq 663  
 QY 199 NDDDFLNYGSSKSI-FAANEK-----GVSWLNYDGTNGEPEPDS 238  
 Db 664 sddnflngltqapffltanggflqsilgtgiry---syevdpdt 708

Search completed: February 26, 2001, 18:58:18  
 Job time: 17650 sec

OM of: US-09-426-072-2\_COPY\_20\_464 to: Issued\_Patents\_NA:\* out\_format : pfs

Date: Feb 27, 2001 2:13 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/6146869/lnat\_13022001\_095136\_14992/app\_query.fasta\_1.1049  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FAPOP=6.000 -FAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000 -USPR=6146869 -CGN1\_1\_73  
-NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-426-072-2\_COPY\_20\_464  
Query length: 445  
Database: Issued\_Patents\_NA:\*  
Database sequences: 280836  
Database length: 80580151  
Search time (sec): 109.120000

score\_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-426-072-1	2347.50	5128.99	3.9e-278	2318	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-350B-1	122.00	213.47	0.0002	4360	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-494-714-1	111.50	191.83	0.0039	3876	1
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US96-10782-1	111.50	191.83	0.0039	3876	1
/cgn2_6/ptodata/2/ina/Backfiles1.seq:5223424-3	109.50	195.55	0.0024	2062	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-119-361-4	104.50	179.03	0.0203	3159	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-336-308A-3	104.50	179.03	0.0203	3159	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-822-324-3	104.50	179.03	0.0203	3159	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-336-308A-9	104.50	168.30	0.0803	7266	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-822-324-5	104.50	168.30	0.0803	7266	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-570-311-28	104.50	166.06	0.1069	8640	1
/cgn2_6/ptodata/2/ina/Backfiles1.seq:5223424-12	103.50	183.93	0.0108	1820	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-579-777A-1	103.50	179.41	0.0193	2585	1
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US95-10509-1	101.50	176.84	0.0268	2241	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-876-280-5	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-049-783-1	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-158-232-5	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-304-626-5	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-316-301A-5	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-611-928-5	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-173-491-5	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-076-137-5	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US92-03624-5	100.50	167.93	0.0842	3771	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-776-597A-1	96.50	162.69	0.1647	2855	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-693-228-1	96.50	162.69	0.1647	2855	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-273-538A-6	96.00	165.00	0.1226	2192	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-949-941B-1	95.50	164.23	0.1353	2136	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-237-919-1	95.00	159.60	0.2449	2808	1
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US95-05518-1	95.00	159.60	0.2449	2808	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-317-785-1	94.50	164.56	0.1297	1755	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-961-083-203	93.50	160.85	0.2087	1972	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-053-614-3	93.50	149.33	0.9147	4821	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-316-397B-3	93.50	149.33	0.9147	4821	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-034-306-3	93.50	149.33	0.9147	4821	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-09-259-437-3	93.50	149.33	0.9147	4821	1
/cgn2_6/ptodata/2/ina/PTCUS_COMB.seq:PCT-US93-09782-3	93.50	149.33	0.9147	4821	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-928-361B-2	93.00	146.50	1.31	5511	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-928-361B-1	93.00	142.82	2.11	7334	1
/cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-425-843-6	92.50	147.58	1.14	4651	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-6	92.50	139.12	3.39	8967	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-9	92.50	139.12	3.39	8967	1

/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-853-659A-64 + 92.50 139.12 3.39 8967  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-853-659A-67 + 92.50 139.12 3.39 8967  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-853-659A-2 + 92.50 126.06 18.07 24701  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-853-659A-3 + 92.50 126.06 18.07 24701

seq\_name: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:US-09-426-072-1

seq\_documentation\_block:

; Sequence 1, Application US/09426072  
; Patent No. 6146869  
; GENERAL INFORMATION:  
; APPLICANT: Kimberly M. Brown  
; TITLE OF INVENTION: Polypeptides Having Phospholipase B  
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same  
; FILE REFERENCE: 5951.000-US  
; CURRENT APPLICATION NUMBER: US/09/426,072  
; CURRENT FILING DATE: 1999-10-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 2318  
; TYPE: DNA  
; ORGANISM: Aspergillus oryzae  
US-09-426-072-1

alignment\_scores:

Quality: 2347.50 Length: 493  
Ratio: 5.275 Gaps: 3  
Percent Similarity: 90.264 Percent Identity: 90.061

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x US-09-426-072-1 ...  
Align seg 1/1 to: US-09-426-072-1 from: 1 to: 2318  
1 SerProValThrSerGluTyrThrSerValArgGluAlaProPheGlyTy 17  
568 AGCCCTGTCACGTCCTGAGTATACGAGCGTCGAGAGCCCTTTCCGATA 617  
17 rlypGroglySerLysGluSerIleGluAsnLeuLysAspLysValGluA 34  
618 CAAAGCTGGCTCCAGAGAGTCCATTGAGAACTTGAAGAGCAAGGTCGAGA 667  
34 snileValTrpLeuIleLeuGluAsnArg ..... 43  
668 ACATTGTCTGGCTTATTCGAGAACAGGTTTGTGTGCCCTTACGATATT 717  
43 ..... 43  
718 CATATGTGGATATATTAATTCCTCANTTCAGCTTGTGTGTGTGAAGACG 767  
44 .....Ser...PheAspAsnIleLeuGlyGlyValar 53  
768 AGCACTAACATATATGTCACAGTCTCTCGATAACATTTCTGGAGAGCGTCG 817  
53 qArgGlnGlyLeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrL 70  
818 CGGCCAAGGACTGGACCAACCCGATCAACACGGCCGCTTGTGCAACTACA 867  
70 ysAsnAlaSerAspProSerSerGlyLysTyrCysThrGlnAlaLysAsp 86  
868 AGAATGGAGGAGCCACCTCTCTGGCAAGTACTGTACTCAGGCCCAAGGAC 917  
87 TyrAspSerValPheAsnAspProAspHisSerValThrGlyAsnAsnLe 103  
918 ATGATATCCGTTGTTCAACCTCCAGACCACTCCGTCGCTGCTGTAATACT 967  
103 uGluPheTyrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerGlyL 120  
968 GGAGTTTCTACGGAATTTACACCCCAACAAATGTTGCGATTGCGCACTGCA 1017

```
120 ysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyr 136
|||||
1018 AGTCTCGCGGACAGTCTGGCTTCTCAACGACAGCTTAACGACTAC 1067
|||||
137 ProLysLeuAlaProGluGluAlaThrArgGlnValMetGlyTyrTyrTh 153
|||||
1068 CCCAAACTGGCCCCCAGAGAGGACAGCAAGCAAGTATGGATATATAC 1117
|||||
153 rGluGluGluValProThrLeuValAspGluPheThrThrP 170
|||||
1118 GGAGGAGGAGGTTCTACGCTCGTGAGCTTGTGGATGAGTCACTACTT 1167
|||||
170 heAsnSerTrpPheSerCysValProGly..... 179
|||||
1168 TCACACAGCTGGTTCTCGTGTTCTCGGCTAAGTCGGGATAAATCTGTAT 1217
|||||
180 .....ProThrAsnProAsnArgLe 186
|||||
1218 ATTTTAAATTTTGACTAACCGGTTCAAAAGGCTTACCAACCCCAACCGCTT 1267
|||||
186 uCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspAsp 203
|||||
1268 GTCCGCTGGCAGGACCGCTGCTGGGCATGGCAGATGACGAGTACT 1317
|||||
203 heLeuAsnTyrGlyLeuSerSerLysSerLysPheGluAlaAlaAsnGlu 219
|||||
1318 TCCTGAACTATGGTATCTCTAGCAAGTCCATCTTCGAGGCGCCCAACGAG 1367
|||||
220 LysGlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluPr 236
|||||
1368 AAGGGGTGCTCTGGCTCACTACCATGGCACCAAGGAGATTCGAACC 1417
|||||
236 oAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSerAsnValV 253
|||||
1418 GGATTCCTCTCTCTACCTACGTCACCAAGAGCTCCCGGTCCAACGTGG 1467
|||||
253 alProValGluAsnPhePheGlnAspAlaTyrLeuGlyValLeuProLys 269
|||||
1468 TGCGCGTGAAGAACTTCTCCAGACGCTACCTCGGTGTCCTCCCTAAA 1517
|||||
270 PheSerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMetH1 286
|||||
1518 TTCTCTTACATTAACCCCTCTCTGTCGGGCACCAACCACTCCATGCA 1567
|||||
286 sProThrGlyAsnValSerTyrGlyGluValPheValLysGlnIleTyrA 303
|||||
1568 CCCACCGGTAACTCTCTACGCTGAGGCTTCTGTCAGCAGATCTATG 1617
|||||
303 spAlaIleArgGlnGlyProGlnTrpAspLysThrLeuPheIleThr 319
|||||
1618 ATGCCATTCGGCAGGGCCCTCAGTGGGACAGACCCCTGCTCTTCATTACC 1667
|||||
320 TyrAspGluThrGlyGlyPheTyrAspHisValProProLeuAlaVa 336
|||||
1668 TAGCAGGAGACCGGTGGCTCTACGACCATGTCCTCCCTCTCGCGGT 1717
|||||
336 larqProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysT 353
|||||
1718 CCGCCCGGCAACCTGACCTACACTGAGACTGGGAGAACGGTCAAGAA 1767
|||||
353 yrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTrpValIle 369
|||||
1768 ACACCTCTTCTCGACCGCTCTGGGTGGCGCATGCGGACCTGGGTTATC 1817
|||||
370 SerProTyrSerLysLysGlyTyrIleGluGlnTyrGlyThrAspProVa 386
|||||
1818 TCCCTTTACATAAGAGGGATACATCGAGAGTACGGAACGGATCCCGT 1867
|||||
386 lThrGlyLysProAlaProTyrSerAlaThrSerValLeuLysThrLeuG 403
|||||
1868 CACGGGCAAGCCGCTCCCTACAGTGCTACCTCCGCTCAAGACTCTCG 1917
|||||
403 lyTyrLeuThrAspIleGluAspPheThrProArgValAlaHisSerPro 419
|||||
```

```
|||||
1918 GATATCTCTGGACATCGAGGACTTCACCCCTCGTGTGCGCCACTCTCCA 1967
|||||
420 SerPheAspHisLeuIleGlyThrThrLeuArgGluAspAlaProIleAl 436
|||||
1968 TCTTTTCGATCACCTGATCGGACGACGCTTTCGTTGAGGATGCTCTCTATTGC 2017
|||||
436 aLeuLysThrProHisThrPheSerVal 445
|||||
2018 TCTCAAGACTCCCATACCTTTTCGGTA 2045
|||||
```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-470-350B-1

```
seq_documentation_block:
; Sequence 1, Application US/08470350B
; Patent No. 5684126
; GENERAL INFORMATION:
; APPLICANT: Li, Xiao
; APPLICANT: Snyder, Solomon H
; TITLE OF INVENTION: Ebnerin: A Secreted von Ebner's Gland
; TITLE OF INVENTION: Protein Associated with Taste Buds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,350B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wolfe, Susan A
; REGISTRATION NUMBER: 33,568
; REFERENCE/DOCKET NUMBER: 01107.48790
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Rattus rattus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 94..3963
; US-08-470-350B-1
```

alignment\_scores:  
Quality: 122.00 Length: 303  
Ratio: 0.878 Gaps: 16  
Percent Similarity: 45.875 Percent Identity: 23.762

alignment\_block:  
US-09-426-072-2\_COPY\_20\_464 x US-08-470-350B-1 ..  
Align seg 1/1 to: US-08-470-350B-1 from: 1 to: 4360

67 CysAsnTyr .....LysAsnAlaSerAspPr 75

```

|||||
2758 TGCACATATGACTATATAGAGATTTTGTATGGCCCCACACAGTTTCACC 2807
      :|||:|||||
75 oSerSerGlyLysTyrCys.....ThrGlnAlaL 85
      :|||:|||||
2808 TCTCATTTGCCCGGTTTGTGATGGGCCATGGGCTCTTTTCATCTTCAACAT 2857
      :|||:|||||
85 yAspTyrAspSerValPheAsnAspProAspHisSerValThr..... 99
      :|||:|||||
2858 CCAACTTCATGTCAGTTGCTTCACCACTGATCAGAGTTTACTCGAGA 2907
      :|||:|||||
100 GlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAsnAsnGlyAlaI 116
      :|||:|||||
2908 GGGTTCGGGCTGACTACTACTACTGAGCTTTGACAAATAATACCAACATCT 2957
      :|||:|||||
116 eAlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnL 133
      :|||:|||||
2958 CCTTTGT.....C 2965
      :|||:|||||
133 euAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnValMet 149
      :|||:|||||
2966 TGTCAATCATGAGAGCCAGTGTGAGCAGGAGCTTACCTTCAGTCCATG 3015
      :|||:|||||
150 GlyTyrTyrThrGluGluValProThrLeuValAspLeuValAspGln 166
      :|||:|||||
3016 GGTACTCTCCAGG.....GATCTTGT..... 3039
      :|||:|||||
166 uphethThrPheAsnSerTrpPheSerCysValProGlyProThrAsnP 183
      :|||:|||||
3040 .ATTCCTGGTTGGAAGCTGAGTTACCACTGTCAGCTCAGATAACACAAA 3088
      :|||:|||||
183 toAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsn 199
      :|||:|||||
3089 GGGAGGTATATTCACAAATCCCTACACAGGCTGGGTTACTACCAACAG 3138
      :|||:|||||
200 ...AspAspPheLeuAsnTyrGlyLeSerSerLysSerIlePheGln 215
      :|||:|||||
3139 GCTCAACAGCAGACCACTCAATC.....TCCAACCTTCCTCAA 3176
      :|||:|||||
215 uAlaAlaAsnGluLysGlyVal..... 222
      :|||:|||||
3177 AGCGCTGTTTCAATGGCATCATCAAAAGGAGAAAGGATCTCCACATCC 3226
      :|||:|||||
223 .....SerTrpLeuAsn.....TyrAsp 228
      :|||:|||||
3227 ATGTCAGCTCAGATGCTTCAGAACACCTGGGTCAACCACTATCATC 3276
      :|||:|||||
229 GlyThrAsn.....GlyGluPheGluPr 236
      :|||:|||||
3277 ACCAACACACAGTCGAGATCCAGGAAGTCCAGTATGGCAATTTGACGT 3326
      :|||:|||||
236 oAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSerAsnVal 253
      :|||:|||||
3327 GAATATTTCTTTTAT.....ACATCTCTCTCTCTCTGT 3361
      :|||:|||||
253 alProValGlu.....AsnPheGlnAspAlaTyr 263
      :|||:|||||
3362 ATCCAGTGCAGCAGCCCATATATGTTGATCTGACCAAGAAATTTGTAC 3411
      :|||:|||||
264 Leu...GlyValLeuProLysPheSerTyrIleAsnProSerCysCysGly 279
      :|||:|||||
3412 CTTGAGGCCGAATCTCTCATCGGATACCTCTTGGCTCTGTTTGGGA 3461
      :|||:|||||
280 .ThrAsnThrAsnSerMetHisProThrGlyAsnValSerTyrGlyGlu 296
      :|||:|||||
3462 CACCTGTGTGGCTTCGCCACATCCCAATGAC..... 3492
      :|||:|||||
296 alPheValLysGlnIleTyrAspAlaIleArgGlnGlyProGlnTrpAsp 312
      :|||:|||||
3493 ..TTCCTGCTTTGACATATGATCTCATCAGAGTGGATGATACGAGAT 3540
      :|||:|||||
313 LysThr 314
      :|||:|||||

```

3541 GAAACT 3546

seq\_name: /cgn2\_6/ptodata/2/ina/5a\_COMB.seq:US-08-494-714-1

```

; seq_documentation_block:
; Sequence 1, Application US/08494714
; Patent No. 5587290
; GENERAL INFORMATION:
; APPLICANT: Klionsky, Daniel
; APPLICANT: Holzer, Helmut
; APPLICANT: Destruelle, Monica
; TITLE OF INVENTION: STRESS TOLERANT YEAST MUTANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,714
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard Aron
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-61036/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8700
; TELEX: 210 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3876 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 579..3701
; US-08-494-714-1

```

## alignment\_scores:

```

Quality: 111.50 Length: 446
Ratio: 0.555 Gaps: 25
Percent Similarity: 45.067 Percent Identity: 19.955

```

## alignment\_block:

```

US-09-426-072-2_COPY_20_464 x US-08-494-714-1
Align seg 1/1 to: US-08-494-714-1 from: 1 to: 3876
3 ValThrSerGluTyr.....ThrSerValArgGluAlaPropheG1 16
||| ||||| :|||:||||| ||| |||:
1914 GTCGATTACCAATACCATATTAACTGTTGATGCTATGCGCTCTCTTTC 1963
16 yTyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValG 33
:||||| :|||:||||| :|||:|||||
1964 CATATACCTCAATGGACACGAGGGATTGATGAC.....GAGTATCTGA 2007
33 luAsnIleValTrpLeuIleLeuGluAsnArgSer..... 44
:||||| :|||:||||| :|||:|||||
2008 GATATACCTACATGGCCAATTATCAAAAACGACGCCCAATTTTACTGCT 2057
45 .....PheAspAsnIleLeuGlyGlyValArgGlnGlyLeuAs 58

```

```

2058 TATGTTAAGTACAAATCTTCCCTAGGATTTGATGAACATATAATTTGAC 2107
      ::::::::::: ||||| ::::
58 pAsnPro.....lleAsnAsnGlyProPheCysAsnT 69
      ::||| ||||| ||||| |||
2108 AGATCCCGCAGAGTTTCCTAATCATCATCAATACGGGGCTTTC..... 2150
69 yrlYsAsnAlaSerAspProSerGlyLysTyrCysThrGlnAlaLys 85
      |||||::: :::: ||::::||:::
2151 ..ACGAATGCTGGCATTTAAACACACTTCTAAAGTGGGCAACAGACATTTGGC 2198
      ::::: ::::: ::::: ||:::
86 AspTyrAspSerValPheAsnAspProAspHisSerValThrGlyAsnAs 102
      ::::: ::::: ||||| ||| :::::
2199 AATCATCTCGCGAGGTCGTTGACCCCAATGGAGTGAATTTCCAAAGA 2248
102 nLeuGluPheTyrGlyThrThrProAsnAsnGlyAlaIleAlaSerG 119
      ::::: ||| ::::: :::::
2249 TATT.....TATATCCCTAGA.....TCCTCAT 2271
119 lYlYsValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAsp 135
      ::::: ::::: |||||
2272 CTACATCATTGGATATATTCTGGT.....ATGAATAGC 2306
136 TyrProLysLeuAlaProGluGluAlaThrArgGlnVal.....MetG 150
      ::::: ::::: |||||
2307 TCAGTGGAGATTAAACAGCGGATGTGACTTAAATGGTTTACCCACTTGG 2356
150 yTyrTyrThrGluGluValProThrLeuValAspLeuValAspGluP 167
      |||| ::::: ||||| :::::
2357 ATATATCAATGATGAATCCATA.....TTGAACATGCANTTAAGATC 2400
167 heThrThrPheAsnSerTrpPheSerCysVal...ProGlyProThrAsn 182
      ::::: ::::: |||||
2401 TTTATTATTATTAGAAAGACAGCTCTGGCTCTGGCGCTGCAATGACATAT 2450
183 ProAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGly..... 197
      ||| ::::: |||||
2451 CCGGTTTTTGTCGCGCAGCTGCTGGT...CTGCTGAATCAGCGCTCTTC 2497
197 ..... 197
2498 TTCTCAAAAGTTACTTATATAATFCGGTCTTCCATCTTACGGGCTCCTT 2547
198 .....LysAsnAspAspAspPheLeuAsnTyrGlyIle 208
      ::::: ||||| ::::: |||||
2548 TCGCTCAATTTAGTGACCAATCAGACGACACTTTTAAACAAACGATTA 2597
209 SerSerLysSerIle...PheGluAlaAlaAsnGluLys..... 220
      ::::: ::::: ||| heGluAlaAlaAsnGluLys
2598 ACCCAGCCAGCATTCCTCTTTTAAACAGCTAACGGTGGATTCTTACAGAG 2647
221 .....GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluP 234
      ||| ::::: |||||
2648 CATTCGTGTTGGTTAACAGGAATCCGATAT.....TCTTATGAGG 2688
234 heGluProAspSerLeuPheThrValAsnGlnThrSerArgSer 250
      ::|||:::
2689 TTGATCCAGTACT.....AAAAAATTAACCGTTGTGTAAGGTTTC 2729
251 AsnValValProValGluAsnPhePheGlnAspAlaTyrLeuGlyVal 267
      ||| ::::: |||||
2730 AAT.....CCAATAGAA.....CTACCGTGTCT 2752
267 uPro.....LysPheSerTyrIleAsnPro..... 275
      ||||
2753 CCCTGGTGATGCTATATAGAACTTCAATATATATGACCCGCTTTTAG 2802
276 .....SerCysCysGlyThrAsnThrAsnSerMetHisProThrGlyAsn 290
      ||| ::::: |||||
2803 ATATAATAATTGACCAACCAATGGTACGATTGTTTCATAAATCAGGAGAT 2852
291 ValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgG 307
      ||| ::::: |||||

```

```

2853 GTT.....CCTATTCATATAAG..... 2870
307 nGlyProGlnTrpAspLysThrLeuPheIleThrTyrAspGluThrG 324
2870 ..... 2870
324 lYgLYpHeTyrAspHisValProProLeuAlaValArgProAspAsn 340
      ::|||
2871 .....ATACCAACACAGATCTCTAATACATGACCAAGAT 2903
341 LeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHis... 356
      ::::: ::::: |||||
2904 ATCACTTCTATATATGTTCCGAAACGAAGAAGAACCAATCTAGAGCG 2953
357 .....PheAspArgLeuGlyArgMetProThrTrpValIleSerP 371
      |||||::: |||
2954 TAGAGACGTCGACCGTGTGGTGATCCCAATG..... 2984
371 rOTyrSerLysLysGlyTyrIleGluGlnTyrGlyThr 383
      ::::: |||||
2985 .....AGGATGGATAGGTATGTACC 3005
seq_name: /cgn2_6/ptodata/2/ina/PCUTUS_COMB.seq:PCT-US96-10782-1
seq_documentation_block:
: Sequence 1, Application PC/TUS9610782
: GENERAL INFORMATION:
: APPLICANT: The Regents of the University
: APPLICANT: of California
: TITLE OF INVENTION: STRESS TOLERANT YEAST MUTANTS
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ROBBINS, BERLINER & CARSON
: STREET: 201 N. Figueroa Street, 5th Floor
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90012-2628
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/10782
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Berliner, Robert
: REGISTRATION NUMBER: 20,121
: REFERENCE/DOCKET NUMBER: 5555-400
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 977-1001
: TELEFAX: (213) 977-1003
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3876 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 579..3701
: PCT-US96-10782-1

```

```

alignment_scores:
  Quality: 111.50      Length: 446
  Ratio: 0.555        Gaps: 25
  Percent Similarity: 45.067  Percent Identity: 19.955

```

## alignment\_block:

```

US-09-426-072-2_COPY_20_464 x PCT-US96-10782-1
Align seg 1/1 to: PCT-US96-10782-1 from: 1 to: 3876

3 ValThrSerGluTyr.....ThrSerValArgGluAlaProPheG1 16
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1914 GTGATTACGAATACCATATTAACGTTGATGCTCGCTATGGCGCTCTTTC 1963

16 yTyrLysProGlySerLysIleGluAsnLeuLysAspLysValG 33
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1964 CATATCTTGAATGGACACGAAGGATTTGATGAC.....GAGTATCTGA 2007

33 LuAsnIleValTrpLeuIleLeuGluAsnArgSer..... 44
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2008 GATATCTTACATGCCAATATATCAAAAACGACGCCCAATTTTACTGCT 2057

45 .....PheAspAsnIleLeuGlyValArgGlnGlyLeuAs 58
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2058 TATGTTAAGTACAAATCTTCCCTAGGATTTGATGAACATATAATTTGAC 2107

58 pAsnPro.....IleAsnAsnGlyProPheCysAsnT 69
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2108 AGATCCCGACGAGTTTGTAAATCAATCAATAACGGGGCTTTC..... 2150

69 yTyrLysAsnAlaSerAspProSerSerGlyLysTyrCysThrClnAlaLys 85
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2151 ..ACGAATGCTGGCATTAAACACCTTCTAAGTGGGACACAGACATTGGC 2198

86 AspTyrAspSerValPheAsnAspProAspHisSerValThrGlyAsnAs 102
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2199 AATCATCTCGCGGAGGCTGTTGACCCCAATGAGTGAATTTCCAAAGA 2248

102 nLeuGluPheTyrGlyThrTyrProAsnAsnGlyAlaIleAlaSerG 119
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2249 TATT.....TATATCCCTAGA.....TCCTCAT 2271

119 lYlsValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAsp 135
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2272 CTACATCACTTTGGAATATCTGCT.....ATGAATAGC 2306

136 TyrProLysLeuAlaProGluAlaThrArgGlnVal.....MetG1 150
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2307 TCAGTGGAGATTAAACGCGGATGACCTTAATGGTTTACCCACTTGG 2356

150 yTyrTyrThrGluGluValProThrLeuValAspLeuValAspGluP 167
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2357 ATATATCAATGATCAATCCATA.....TTGAACAATGCAATTAAGATC 2400

167 heThrThrPheAsnSerTrpPheSerCysVal....ProGlyProThrAsn 182
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2401 TTTATTATTATTCAGAAAGACAGTCTGCGCTGGGCTGCAATGACATAT 2450

183 ProAsnArgLeuCysAlaLeuAlaGlyThrAlaIleGlyHisGly..... 197
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2451 CGGGTTTTTGTGGCCGACCTGCTGGT...CTGCTGAATCAGCGCTCTTC 2497

197 ..... 197

2498 TTCCTCAAGTTACTTATATAATCGGTTCTTCCATACTTACGGGCTCCTT 2547

198 .....LysAsnAspAspPheLeuAsnTyrGlyIle 208
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2548 TCGCTCAATTTAGTAGCAATCAGACGACCACTTTTAAACAAACGGATTA 2597

209 SerSerLysSerIle....PheGluAlaAlaAsnGluLys..... 220
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2598 ACCCAGCCACCATTCCTTTTAAACAGCTAACCGTGGATTTCTACAGAG 2647

221 .....GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluP 234
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2648 CATTCCTGTTGGGTTAACAGGAATCCGATAT.....TCTTATCAGG 2688

```

```

234 heGluProAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSer 250
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2689 TTGATCCAGTACT.....AAAAAATTAAACCGTTTGTAAAGTTC 2729

251 AsnValValProValGluAsnPhePheGlnAspAlaTyrLeuGlyValle 267
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2730 AAT.....CCAATAGAA.....CTACCGTTGCT 2752

267 uPro.....LysPheSerTyrIleAsnPro..... 275
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2753 CCTCGGTGGTATCGCTATTAGAACTTCAATATATGAACCCAGTTTAG 2802

276 .....SerCysCysGlyThrAsnThrAsnSerMetHisProThrGlyAsn 290
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2803 ATATAATAATGAGCACCAATGTCAGGATTTGTCATAAATCAGGAGAT 2852

291 ValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgG1 307
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2853 GTT.....CCTATTCAATAAAG..... 2870

307 nGlyProGlnTrpAspLysThrLeuPheIleThrTyrAspGluThrG 324
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2870 ..... 2870

324 lYglyPheTyrAspHisValProProLeuAlaValArgProAspAsn 340
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2871 .....ATACCAACACAGATCTCTAATACATGACCAAGAT 2903

341 LeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHis... 356
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2904 ATCACTTCATATAATGTTCCGAAACGAAAGAAACCAATCTAGACGG 2953

357 .....PheAspArgLeuGlyGlyArgMetProThrTrpValIleSerP 371
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2954 TAGAGACGTCGACCGTGTGGTGATCCAATG..... 2984

371 roTyrSerLysLysGlyTyrIleGluGlnTyrGlyThr 383
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2985 .....AGATGGATAGGTATGGTACC 3005

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5223424-3
seq_documentation_block:
; Patent No. 5223424
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,
; RICHARD D.
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
; HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
; ACID SEQUENCE
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/225,032
; FILING DATE: 27-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 78,519
; FILING DATE: 27-JUL-1987
; APPLICATION NUMBER: 933,107
; FILING DATE: 20-NOV-1986
; APPLICATION NUMBER: 902,887
; FILING DATE: 02-SEP-1986
; APPLICATION NUMBER: 887,140
; FILING DATE: 17-JUL-1986
; APPLICATION NUMBER: 823,102
; FILING DATE: 27-JAN-1986
; APPLICATION NUMBER: 773,430
; FILING DATE: 06-SEP-1985
; SEQ ID NO:3:
; LENGTH: 2062
5223424-3

```

```

alignment_scores:
Quality: 109.50 Length: 402

```

Ratio: 0.622  
Percent Similarity: 43.781 Gaps: 22  
Percent Identity: 21.393

## alignment block:

US-09-426-072-2\_COPY\_20\_464 x 5223424-3

Align seg 1/1 to: 5223424-3 from: 1 to: 2062

```

91  PheAsnAspProAspHisSerValThrGlyAsnAsnLeu..... 103
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
627  TTCAATCCAGCGGACTGGCAGTTAATATCAACAACATGACAGAAATAAA 676
104  .....GluPheTyr..... 106
677  CTTAGTTAGTTTGAACAAGAATATTCATGCTAGTACTTAAACAATA 726
107  .....GlyThrTyrThrPro.....AsnAsnGlyAlaIle 116
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
727  CAGAATCAGCAACCTCCACCACCATCAAAATATATATATGATCTAACT 776
117  AlasrGlyLysValAlaAspGlnSerGlyPheLeuAsnAlaGlnLe 133
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
777  GCAAGCTTAATGGTCGACCTAGACACCAATTAACACACTT..... 815
133  uAsnAspTyrProLysLeuAlaProGluAlaThrArgGlnValMetG 150
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
816  ....CCATACACACCGACGACGACCTAGAAGTGAACA.....CTTG 852
150  lyTyrTyrThrGluGluValProThrLeuValAspLeuValAspGlu 166
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
853  GTTTTAT.....CCATGGTTACTTACA.....AAA 878
167  PheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsn 183
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
879  CCAACTCAATACAGATATTACCTATCATGCATCAGAAACCTAAATCCACC 928
183  oAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsn 200
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
929  A.....ACATACACTGGACATCAACACAAA 954
200  spAspAspPheLeuAsnTyrGlyLeSerSerLysSerIlePheGluAla 216
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
955  TAACAGACTCAATACAAACAGAGCTACACAGTACATATGTCTACACA 1004
217  AlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGly.....Th 230
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1005  ATAGAAATGCAGTACCAATTCATCTTCTAAGAACTGGAGATGAATTC 1054
230  rAsnGly.....GluPheGluProAspSerLeuPhePheThrTyrVal 245
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1055  CACAGGAATATATCACTTTTGACACAAAACCATTAATAATTAACCTCAT 1104
245  snGlnThrSerArgSerAsnValIleProValGluAsnPheGln... 260
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1105  GGCAACAAACAGATCTCTAGGACTGCCTCCAAAAGTACTAAGTGAACCT 1154
261  .....AspAlaTyrLeuGlyValLeuProLysPheSerTyrIle 273
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1155  ACCACAGAAAGGAGACCAACACCCAGGAACACTACCA..... 1190
273  eAsnProSerCysGlyThrAsnThr...AsnSerMetHisProThrG 289
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1191  .....GGACTAACACAGAAAGGTTATCACCACAA 1224
289  lyAsnValSerTyrGlyValPheValLysGlnIleTyrAspAlaIle 305
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1225  TTAATAATAGCTACACAGAAGCA.....ACAGCACTT 1256
306  ArgGlnGlyPro...GlnTrpAspLysThrLeuLeuPheIleThrTyr 321
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1257  AGCCAGCTCAGTAGGATATATACCATCATGATTTTCACTACTC 1306
321  pGluThrGlyGlyPheTyrAspHisValProProLeuAlaValArgp 338

```

```

1307  CAATGGTGACCAATTTCTA.....ACTCTATAGTACCACAG 1344
1338  roAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThr 354
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1345  CAGACACACAATATTATGATGATGAACCAATGTGTCTATAGATTACA 1394
355  Leu.....HisPheAspArgLeuGlyGlyArgMetProTh 366
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1395  ATGGGTTTACCACATGGACATTAACCCACATCTTCAACAGAGCTAGAAG 1444
366  rtrpValIleSerProTyrSerLysLysGlyTyrIle.....GluGln 381
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1445  ATACACATCAATCCACAAGTAATGTGGAGAGCTCCAAAGCAACAT 1494
381  yGly..... 382
1495  TTAATCAACAGGCACCACCTAAACCTAGAAATAAATAAATAATGGAACACTT 1544
383  .....ThrAspProValThrGlyLysPro..... 390
1545  TTACCTTCAGATCCATAGGAGGAAATCTAACAGCAATTTTCATGAATAC 1594
391  .....AlaProTyrSerAla.....ThrSerValLeuL 400
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1595  ACTCAATACATATGGACCATTAACAGCACCTAAACAATACTGCACCTGTAT 1644
400  yThrLeuGlyTyrLeuTrpAspIleGlu.....AspPheThrPro 413
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1645  TTCCAATGGTCAATATGGATTAAGAACTTGATACAGATCTAAACCT 1694
414  ArgVal 415
1695  AGACTA 1700

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-119-361-4

## seq\_documentation\_block:

```

; Sequence 4, Application US/0819361
; Patent No. 523390
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis Arginine-specific Proteinase
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,361
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 21-93
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-499-8080
; TELEFAX: 303-499-8089
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3159 base pairs

```



```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 949..3159
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1630..3105
;
us-08-119-361-4

```

## alignment\_scores:

```

    Quality: 104.50      Length: 521
    Ratio: 0.488        Gaps: 30
    Percent Similarity: 41.075      Percent Identity: 19.962

```

## alignment\_block:

```

us-09-426-072-2_copy_20_464 x us-08-119-361-4

```

```

Align seg 1/1 to: us-08-119-361-4 from: 1 to: 3159

```

```

17 TyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValGI 33
||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1615 TACGAGCGGGGGCGTTACACACCGGTAGAG.....GAAAAACA 1652

33 uasnIleValTrpLeuIleLeuGluAsnArgSerPheAspAsnIleLeuG 50
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1653 AAATGCTGATGATCGTCATCGTAGCCAAAGATATCAG.....G 1693

50 lYgLyValArgArgGlnGlyLeuAspAsnProIleAsnAsnGlyProPhe 66
|| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1694 CAGATATTTAAA.....GATTTC 1710

67 CysAsnTyrIleAsnAlaSerAspProSerSerGlyLysTyrCysThrGI 83
||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1711 GTTGATTGGAACCAACACGCGGT.....CTCCGTACCGA 1745

83 nAlaLysAspTyrAspSerValPheAsnAspProAspHisSerValThrG 100
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1746 GGTGAA.....GTGGCAGAGATATGCTCTCCCGTTACAG 1783

100 lYAsnAsnLeuGluPheTyr.....GlyThrTyrThrProAsnAsnGly 114
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1784 CTAATGCTATTACGAGTTCGTTAAGCAAGAATACGAGAAAGAGTAAT 1833

115 AlaIleAlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAl 131
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1834 GATTGACCTATGCTTTTGGTTGGCGATCAG..... 1866

131 sGluLeuAsnAspTyrPro.....LysLeuAlaProGluAlaThrArg 147
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1867 .....AAAGATATTCTCGCAAAATTTACTCCGGGATCAAAATCCGACC 1909

147 InValMetGly.....TyrTyrThrGlu..... 154
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1910 AGGTATATGACAATAGTAGTAATGACCACCTACACGAAGTCTTCATC 1959

155 .....GluGluValProThrLeuValAspLe 163
||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1960 GGTGTTTCTCATGTGAGACCAAGAGGATCTGAAGACACAAATCGATCG 2009

163 uValAspGluPhe.....ThrThrPheAsnSerTrpPheSerC 176
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2010 GACTATTCACTATGCGCAATATAACCGAGGACAAATGGCTCGGT. 2058

176 ysValProGlyProThrAsnProAsnArgLeuCys.....AlaLeu 189
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2059 .....CAGGCTCTTTGATTGCTTCGGCTGAA 2085

```

```

190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTyr 206
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2086 GGAGGCCATCCGCGACACAATGTTGAAGTGATATCCAGCATGAGAATGT 2135

206 r..... 206

2136 AATCGCCAATCTGCTTACCCAGTATGCTATACCAAGATTATCAATGT 2185

207 .....GlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLys 220
||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2186 ATGATCGGGAGTAACTCTTAAAAACAATTATTGATGCTTCAAC...GGA 2232

221 GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAs 237
||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2233 GGAATCTCGTTGTCACATATACGGGCCAGCGTAGCAACACAGCTTGGGG 2282

237 pSerLeuPhePhe.....ThrTyrValAsnGlnThrSerArgSerAsnV 252
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2283 TACGCTCCTACCTTCGGCACCACTCATGTGAAGCAGCTTACCAACAGCAACC 2332

252 alValProValGluAsnPhePheGlnAspAla..... 262
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2333 AGCTACCG.....TTTATTTTCGACGTAGCTTGTGTGAATGGCGAT 2373

263 TyrLeuGlyValLeuProLysPhe..... 270
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2374 TTCCTATTACGATGCTTTCGTCGAGAGCCCTGATGCGTGACACAAA 2423

271 .....SerTyrIleAsnP 275
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2424 AGATGTGAAGCCGACAGGACTGTGTGTCATCATAGCGCTACGATCAACC 2473

275 roSer.....Cys 277
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2474 AGCTTGGGCTTCTCTATGCGCGGCGAGTGAAGATGAACGAATTCGTG 2523

278 CysGlyThrAsnThrAsnSerMetHisProThr...GlyAsnValSerTy 293
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2524 TCGAAACACCCGACCAACATCAAGCGTACTTTCGGTGGTGTCCACCAT 2573

293 rGlyGluValPheVal.....LysGlnIleTyrAspAlaIleArg 307
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2574 GAACGCTATGCTTGTCTATGCTGGAAGATATAAAGAGGTGGTGAGAAGA 2623

307 InGlyProGlnTrp.....AspLysThrLeuLeuPheIleThr 319
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2624 TGCTCGACACATGGAGCTGTTTCGGCGACCCCTCGCTGCTGCTGTAC 2673

320 TyrAspGluThrGlyGlyPheTyrAspHisValProProLeuAlaVal 336
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2674 CTT.....GTCCCGACCAAAATGCAGGT 2696

336 L....ArgProAsnLeuThrTyrThrGluThr..... 346
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2697 TACGGCTCCGGCTCAGATTAATTTGACGGATGCTTCAGTCAACGTATCTT 2746

347 .....AlaLysAsnGlyGlnLysTyr 353
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2747 GCGATTATANTGGCTATTGCTACCATTTTCAGCCCAATGGAAGATGTC 2796

354 .....ThrLeuHisPheAspArgLe 360
||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2797 GGTTCGCTGAGTGTGAAANTGGAACAGCTACAATCAATCTGCAGAGTCT 2846

360 uGlyGlyArgMetPro...ThrTrpValIleSerProTyrSerLysLysG 376
|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2847 GACAAATGAAGCAGCGCTTACCCTTACAGTAGTTGTTACAACAAACAGA 2896

376 lYThrIleGluGlnTyrGlyThrAspProValThrGlyLysProAlaPro 392
||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2897 CGGTATTAAAGACCATCAACACTAAT.....GGTAGCGCTTAACCCC 2937

393 Tyr.....SerAlaThrSerValLeuLysThrLeuGI 403

```

```

|||||
2938 TACCAGCCGTTTCCAACTTGACAGCTACACGCGGTGAGAAAGTAAC 2987

```

```

403 yTyrLeuTrpAsp 407

```

```

2988 GCTCAAGTGGGAT 3000

```

```

seq_name: /cgn2_6/ptodata/2/lna/6_COMB.seq:US-08-336-308A-3

```

```

seq_documentation_block:
; Sequence 3, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3159 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 949..3159
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1630..3105
US-08-336-308A-3

```

```

alignment_scores:
  Quality: 104.50
  Ratio: 0.488
  Gaps: 30
  Percent Similarity: 41.075
  Percent Identity: 19.962

```

```

alignment_block:

```

```

US-09-426-072-2_COPY_20_464 x US-08-336-308A-3
Align seg 1/1 to: US-08-336-308A-3 from: 1 to: 3159
17 TyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValG1 33
|||||
1615 TACGAGCGGGGGTTCACACCGGTAGAG.....GAAAAACA 1652
33 uAsnIleValTrpLeuLeuGluAsnArgSerPheAspAsnIleLeuG 50
|||||
1653 AAATGGTGGTATGCTGCTATCGTAGCCAAAAGATATGAG.....G 1693
50 IyGlyValArgArgGlnGlyLeuAspAsnProIleAsnAsnGlyProPhe 66
|||
1694 GAGATATTAA.....GATTTC 1710
67 CysAsnTyrLysAsnAlaSerAspProSerSerGlyLysTyrCysThrG1 83
|||||
1711 GTTGATTGGAAACCAACCGCGGT.....CTCCGTACCGA 1745
83 nAlaLysAspTyrAspSerValPheAsnAspProAspHisSerValThrG 100
|||||
1746 GGTGAAA.....GTGCGAGAAGATATTGCTTCCTCCGTTACAG 1783
100 IyAsnAsnLeuGluPheTyr.....GlyThrTyrThrProAsnAsnGly 114
|||||
1784 CTAATGCTATTCAGCAGTTCGTTRAGCAAGAATACGAGAAAGAGTAAT 1833
115 AlaIleAlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAl 131
|||||
1834 GATTGACCTATGTTCTTTGGTTGGCGATCAC..... 1866
131 aglnLeuAsnAspTyrPro...LysLeuAlaProGluGluAlaThrArgG 147
|||||
1867 .....AAAGATATTCCTGCCAAATTAATCTCGGGGATCAATCCGACC 1909
147 InValMetGly.....TyrTyrThrGlu..... 154
|||||
1910 AGGTATATGGACAAATAGTAGTAAATGACCCTACACGAAGTCTTCATC 1959
155 .....GluGluValProThrLeuValAspLe 163
|||||
1960 GGTCTGTTCTCATGTGAGAGCAAGAGGATCTGAAGACACAAATCGATCG 2009
163 vValAsnGluPhe.....ThrThrPheAsnSerTrpPheSerC 176
|||||
2010 GATATTCACTATGAGCGCATATATACCGGAGACAAATGGCTCGGT. 2058
176 ysValProGlyProThrAsnProAsnArgLeuCys.....AlaLeu 189
|||||
2059 .....CAGGCTCTTTGATTGTTCTTCGCTGCTGAA 2085
190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206
|||||
2086 GGAGGCCCATCCGACACAAATGGTGAAGTGATATCCAGCATGAGATGT 2135
206 r..... 206
2136 AATCGCCAATCTGCTTACCAGTATGGCTATACCAAGATTATCAATGTT 2185
207 .....GlyIleSerLysSerIlePheGluAlaAlaAsnGluLys 220
|||||
2186 ATGATCCGGGAGTACTCTCTAAACATATATGATGCTTTCAAC...GGA 2232
221 GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAs 237
|||||
2233 GGAATCTCGTTGTTCAACTATACGGGCCACGGTAGCGAAACACAGCTTGGG 2282
237 pSerLeuPhePhe.....ThrTyrValAsnGlnThrSerArgSerAsnV 252
|||||
2283 TACGCTCACTTCGGCACCACCTCATGTGAGAGAGCGCTTACCACAGCACC 2332
252 alValProValGluAsnPheGlnAspAla..... 262

```

```

2333 ACCTACCG.....TTTATTTTCAGCTAGCTTGTGTGAATGGCGAT 2373
263 TyrLeuGlyValLeuProLysPhe..... 270
2374 TTCCCTATTTCAGCATGCTTTCGTCGACAGGCCCTGATGCGTGCACAAAA 2423
271 .....SerTyrIleAsn 275
2424 ACATGGTAAGCCGACAGGTACTTGTGTATCATAGCGTCTACGATCAACC 2473
275 rOser.....Cys 277
2474 ACTCTTGGGCTTCTCTATGCGCGGCGAGGATGAGATGAACAAATTCG 2523
278 CysGlyThrAsnThrAsnSerMetHisProThr.....GlyAsnValSerTy 293
2524 TCGGAAACACCCGACCAACATCAACGGTACTTTCGGTGGTGCACCAT 2573
293 rGlyGluValPheVal.....LysGlnIleTyrAspAlaIleArg 307
2574 GAACGGTATGTTGCTATGCTGTAAGAAATATATAAAAGGATGCTGAGAAGA 2623
307 lnglyProGlnTTP.....AspLysThrLeuLeuPheIleThr 319
2624 TCGTCGACATGGACTGTTTCGGCGACCCCTCGCTGCTGCTCGTGACA 2673
320 TyrAspGluThrGlyGlyPheTyrAspHisValProProLeuAlaIva 336
2674 CTT.....GTCCGACCAAAATGCAGGT 2696
336 L.....ArgProAsnLeuThrTyrThrGluThr..... 346
2697 TACGCTCGGCTCAGATTAAATTCAGCGGATGCTTCAGTCACGATCTT 2746
347 .....AlaLysAsnGlyGlnLysTyr 353
2747 GCGATTATAAGTGTGCTATTGCTACCATTTTCAGCCAAATGGAAGATGTC 2796
354 .....ThrLeuHisPheAspArgLe 360
2797 GCTTTCGCGCTTTCGAAATGCAACAGCTACAAATCAATCTGCACAGTCT 2846
360 uGlyGlyArgMetPro...ThrTrpValIleSerProTyrSerLysLys 376
2847 GACAAATGAAGACGCGCTTACCCCTTACAGTAGTTGTTACACAAAGAGA 2896
376 LYTrIleGluGlnTyrGlyThrAspProValThrGlyLysProAlaPro 392
2897 CGGTTATTAGACCATCAACACTAAT.....GGTGAGCCTAACCC 2937
393 Tyr.....SerAlaThrSerValLeuLysThrLeuG 403
2938 TACGAGCCGCTTTCACAACTTGACAGCTACAAACGAGGTCAGAAAGTAAC 2987
403 YTrIleuTriPasp 407
2988 GCTCAAGTGGGAT 3000

```

seq\_name: /cgn2\_5/ptodata/2/ina/6\_COMB.seq:us-08-822-324-3

```

seq_documentation_block:
; Sequence 3, Application US/08822324
; Patent No. 6129917
; GENERAL INFORMATION:
; APPLICANT: Potempa, Jan S.
; APPLICANT: Travis, James
; APPLICANT: Genco, Caroline A.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,945
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 103-95 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 488-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3159 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Porphyromonas gingivalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 949..3159
; US-08-822-324-3

```

```

alignment_scores:
  Quality: 104.50      Length: 521
  Ratio: 0.488        Gaps: 30
  Percent Similarity: 41.075  Percent Identity: 19.962

```

```

alignment_block:
US-09-426-072-2_COPY_20_464 x US-08-822-324-3 ..
Align seg 1/1 to: US-08-822-324-3 from: 1 to: 3159
17 TyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValG 33
|||||
1615 TACGAGCGGCGGCTTACACACCGGTAG.....GAAACACA 1652
33 uAsnIleValTrpLeuIleLeuGluAsnArgSerPheAspAsnIleLeu 50
|||||
1653 AAATGGTGTATGATGCTATCTAGTACCAAAAGTATGAG.....G 1693
50 lyGlyValArgArgGlnGlyLeuAspAsnProIleAsnAsnGlyProPhe 66
||
1694 GAGATATTAAA.....GATTTC 1710
67 CysAsnTyrLysAsnAlaSerAspProSerSerGlyLysTyrCysThrG 83
|||||
1711 GTTGATTGGAAAAACCAACCGGT.....CTCCGTACCGGA 1745
83 nAlaLysAspTyrAspSerValPheAsnAspProAspHisSerValThr 100
|||||
1746 GTGAAA.....GTGCAAGAGATATTGCTTCTCCCGTTACAG 1783
100 lyAsnAsnLeuGluPheTyr.....GlyThrTyrThrProAsnAsnGly 114

```

```

1784 CTAATGCTATTGAGCAGTTCGTTAAGCAAGAATACGAGAAAGAGTAAT 1833
115 AlaIleAlaSerGlyValValAlaAlaAspGlnSerGlyPheLeuAsnAl 131
1834 GATTGACCTATGTTCTTTGGTTGGCGATCAC..... 1866
131 aGlnLeuAsnAspTyrPro...LysLeuAlaProGluGluAlaThrArgG 147
1867 .....AAAGATATTCTCCCAAAATTACTCCGGGGATCAAAATCCGACC 1909
147 InValMetGly.....TyrTyrThrGlu..... 154
1910 AGGTATATGCAAAATAGTAGTAATGACCACCTACACGAAGTCTTCATC 1959
155 .....GluGluValProThrLeuValAspLe 163
1960 GGTGCTTTCTCATGTGAGACAAAGAGGATCTGAAGACACAAATCGATCG 2009
163 uValAspGluPhe.....ThrThrPheAsnSerTrpPheSerC 176
2010 GACTATTACTATGACGCAATATATACACGCAAGACAAATSGCTCGGT. 2058
176 ysValProGlyProThrAsnProAsnArgLeuCys.....AlaLeu 189
2059 .....CAGGCTCTTTGTATTGCTTGGGCTGAA 2085
190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206
2086 GGAGGCCCATCGCAGACAATGCTGAAGTGATATCCAGCATGAGAATGT 2135
206 r..... 206
2136 AATCGGCAATCTGTTACCCAGTATGGCTATACCAAGATTATCAAAATGTT 2185
207 .....GlyIleSerSerIlePheGluAlaAlaAsnGluLys 220
2186 ATGATCGGGAGTAACCTCTAAACACATTATTGATGCTTTCAC...GGA 2232
221 GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAs 237
2233 GGAATCTGTTGGTCACTATACGGGCCAGCGTAGCGGAACACAGCTTGGGG 2282
237 pSerLeuPhePhe.....ThrTyrValAsnGlnThrSerArgSerAsnV 252
2283 TAGCTCTCACTTCGGCACCACTCATCTGAAGCAGCTTACCAACAGCAACC 2332
252 alValProValGluAsnPhePheGlnAspAla..... 262
2333 AGTACCG.....TTATTTCGACGTAGCTTGTGTGAATGGCGAT 2373
263 TyrLeuGlyValLeuProLysPhe..... 270
2374 TTCCTATTGAGCATGCTTCTTCGAGAGCCCTGATGGTGACAAAA 2423
271 .....SerTyrIleAsnP 275
2424 AGATGTAAGCGACAGGTACTTCTCTATCATAGCGTCTACCATCAACC 2473
275 roSer.....Cys 277
2474 AGTCTTGGGCTTCTCTATCGCGGCGAGGATGAGATGAACGAATTTCTG 2523
278 CysGlyThrAsnThrAsnSerMetHisProThr...GlyAsnValSerTy 293
2524 TGGAAAAACCCGACACACATCAAGCGTACTTTCGGTGGTGTCCACCAT 2573
293 rglyGluValPheVal.....LysGlnIleTyrAspAlaIleArgG 307
2574 GAACGGTATGTTGCTATGGTGAAAGATATAAAAGATGGTGAAGA 2623
307 InGlyProGlnTrp.....AspLysThrLeuLeuPheIleThr 319
:: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

2624 TGCTGCAACATGGACTGTTTTCGGCGACCCCTCGCTGCTGCTCGTACA 2673
320 TyrAspGluThrGlyGlyPheTyrAspHisValProProProLeuAlaVa 336
2674 CTT.....GTCCGACCAAAATCCAGGT 2696
336 l...ArgProAspAsnLeuThrTyrThrGluThr..... 346
2697 TAGGCTCCGCTCAGATTAATTTGACGGATGCTTCAGTCAACAGTATCTT 2746
347 .....AlaLysAsnGlyGlnLysTyr 353
2747 GCGATTATAATGGTCTATTGCTTACCATTTCCAGCAATGGAAGATGTTT 2796
354 .....ThrLeuHisPheAspArgLe 360
2797 GGTTCCTGCAGTTGTCGAAATGSAACAGCTACAATCAATCTGCACAGTCT 2846
360 uGlyGlyArgMetPro...ThrTrpValIleSerProTyrSerLysLysG 376
2847 GACAAATGAAGCAGCGTTACCTTACAGTAGTTGGTTACAACAAGAGA 2896
376 lyTyrIleGluGlnTyrGlyThrAspProValThrGlyLysProAlaPro 392
2897 CGTTATTAAAGACCATCAACACTAAT.....GGTGAGCCTAACCC 2937
393 Tyr.....SerAlaThrSerValLeuLysThrLeuG1 403
2938 TACGACCCCGTTTCCAACTTGACAGCTACAACGCGAGGTGAGAAAGTAAC 2987
403 yTyrLeuTrpasp 407
2988 GCTCAAGTGGGAT 3000

seq.name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-336-308A-9

seq documentation block:
; Sequence 9, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 949..6063
; US-08-336-308A-9

alignment_scores:
    Quality: 104.50      Length: 521
    Ratio: 0.488        Gaps: 30
    Percent Similarity: 41.075      Percent Identity: 19.962

alignment_block:
US-09-426-072-2_copy_20_464 x US-08-336-308A-9 ..
Align seg 1/1 to: US-08-336-308A-9 from: 1 to: 7266

17 TyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValG1 33
|||||
1615 TAGAGCGGGGGGTTTACACACCGGTAGG.....GAAAAACA 1652

33 uAsnIleValTrpLeuIleGluAsnArgSerPheAspAsnIleLeuG 50
|||||
1653 AAATGGTGTATGCTATCGTATCGTAGCCAAAAGATATGAG.....G 1693

50 lYcIyValArgArgGlnGlnGlyLeuAspAsnProIleAsnAsnGlyProPhe 66
|||
1694 GAGATATTAA.....GATTTC 1710

67 CysAsnTyrLysAlaSerAspProSerSerGlyLysTyrCysThrG1 83
|||||
1711 GTTGATTGGAAAAACCAACCGGT.....CTCCGTACCGA 1745

83 nAlaLysAspTyrAspSerValPheAsnAspProAspHisSerValThrG 100
|||||
1746 GGTGAAA.....GTGGCAGAGAGATATTGCTCTCCCGTTACAG 1783

100 lYAsnAsnLeuGluPheTyr.....GlyThrTyrThrProAsnAsnGly 114
|||||
1784 CTAATGCTATTACAGAGTTCGTTAAGCAAGAATAACGAGAAAGGTAAT 1833

115 AlAlleAlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAl 131
|||||
1834 GATTTGACCTATGTTCTTTGGTGGCGATCAC..... 1866

131 agInLeuAsnAspTyrPro...LysLeuAlaProGluGluAlaThrArgG 147
|||||
1867 .....AAGATATTCTGCCAANAATTACTCCGGGATCAANTCCGACC 1909

147 InValMetGly.....TyrTyrThrGlu..... 154
|||||
1910 AGGTATATGGACAAATAGTAGTAAATGACCACATACAAACGAAGTCTTCATC 1959

155 .....GluGluValProThrLeuValAspLe 163
|||||
1960 GCTCGTTTTCATGTCGAGAGCAAGAGGATCTCAACACACAAATCGATCG 2009

163 uValAspGluPhe.....ThrThrPheAsnSerTrpPheSerC 176
|||||
2010 GACTATTCACTATGAGCGCAATATAACCGGAGCAACAATGGTTCGGT. 2058

176 ysValProGlyProThrAsnProAsnArgLeuCys.....AlaLeu 189
|||||

```

```

2059 ..... |||||
CAGGCTCTTTGTATTCTCTTCGGCTGAA 2085

190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206
|||||
2086 GGAGGCCATCCGCAGACAAATGGTGAAGTGATATCCAGCATGAGAATGT 2135

206 r..... 206

2136 AATCCCAANTCTGCTTACCCAGTAGTGGCTATACCAAGATTATCAAAATGTT 2185

207 .....GlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLys 220
|||||
2186 ATGATCCGGGAGTAGTCTCTTAAAAACATTATTGATGCTTTCAAC...GGA 2232

221 GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAs 237
|||||
2233 GGAATCTCGTTGGTCAACTATACGGCCACGGTAGGCAAAACAGCTTGGG 2282

237 pSerLeuPhePhe.....ThrTyrValAsnGlnThrSerArgSerAsnV 252
|||||
2283 TAGCTCTCACTTCGGCACCACCTCATGTGAAGCAGCTTACCACAGCAACC 2332

252 alValProValGluAsnPheGlnAspAla..... 262
|||||
2333 AGCTACCG.....TTTATTTTCGACGTAGCTTCTGTGAATGGCGAT 2373

263 TyrLeuGlyValLeuProLysPhe..... 270
|||||
2374 TTCTTATTACGATGCTTGTCTTCGAGAACCCCTGATCGTGCACAAAA 2423

271 .....SerTyrIleAsnP 275
|||||
2424 AGATGGTAAGCCAGCAGGTACTGTGTGTATCATAGCTCTACGATCAACC 2473

275 rSer.....Cys 277
|||||
2474 AGTCTTGGGCTTCTCTATCGCGGGCAGGATGAGATGAACGAAATTCG 2523

278 CysGlyThrAsnThrAsnSerMetHisProThr...GlyAsnValSerTy 293
|||||
2524 TCGGAAACACCCGAAACACATCAAGCGTACTTTCGGTGGTGTCACCAT 2573

293 rGlyGluValPheVal.....LysGlnIleTyrAspAlaIleArgG 307
|||||
2574 GAACGCTATGTTGCTATGTTGGAAGATATAAAGAGATGGTGAGAAGA 2623

307 lInGlyProGlnTrp.....AspLysThrLeuLeuPheIleThr 319
|||||
2624 TCCTGCACACATGGACTGTTTCGGCGACCCCTCGCTCGTCTCGTAC 2673

320 TyrAspGluThrGlyGlyPheTyrAspHisValProProProLeuAlaVa 336
|||||
2674 CTT.....GTCCCGACCAAAATGCAGGT 2696

336 l...ArgProAspAsnLeuThrTyrGluThr..... 346
|||||
2697 TACGCTCCGCTCAGATTAAATTCACGGATGCTTCAGTCAACGTATCTT 2746

347 .....AlaLysAsnGlyGlnLysTyr 353
|||||
2747 GCGATTATAATGGTGTATTGCTACCATTTTCAGCCAATGGAAGATGTTTC 2796

354 .....ThrLeuHisPheAspArgLe 360
|||||
2797 GGTCTCCAGTTGCGAAAATGGAACAGCTACAAATCAATCTGCACAGGTCT 2846

360 uGlyGlyArgMetPro...ThrTrpValIleSerProTyrSerLysLysG 376
|||||
2847 GACAATGAAGACGACGCTTACCCCTTACAGTAGTTGGTTACACAAAGAGA 2896

376 lYTyrlleGluGlnTyrGlyThrAspProValThrGlyLysProAlaPro 392
|||||

```



```

263 TyrLeuGlyValLeuProLysphe..... 270
2374 TTCCTATTACGATGCTTTCCTTCGACAGAGCCCTGATCGGTGCACAAA 2423
271 .....SertYrileAsnP 275
2424 AGATGGTAAGCCGACAGGTACTTGTCTATCATAGCGTCTACGATCAACC 2473
275 roSer.....Cys 277
2474 AGTCCTGGCTTCCTCTATGCCGGCAGAGTACGATGAACGAAATTCGT 2523
278 CysGlyThrAsnThrAsnSerMethHisProThr....GlyAsnValSerTy 293
2524 TCGGAAACACCCGACCAACATCAAGCGTACTTTCGTGGTGTCCACCAT 2573
293 rGlyGluValPheVal.....LysGlnIleTyrAspAlaIleArgG 307
2574 GAACGGTATGTTGCTATGCTGCGAAAGTATAAAAGGATGCTGAGAAGA 2623
307 InGlyProGlnTrp.....AspLysThrLeuLeuPheIleThr 319
2624 TGCTCGACATGACGTGTTTTCGGCGACCCCTCGCTGCTGCTGCTACA 2673
320 TyrAspGluThrGlyGlyPheTyrAspHisValProProLeuAlaIle 336
2674 CTT.....GTCCCGACCCAAATGTCAGGT 2696
336 I....ArgProAspAsnLeuThrTyrGluThr..... 346
2697 TACGGCTCCGGCTCAGATTAAATTTCAGCGATGCTTCAGTCAACGTATCTT 2746
347 .....AlaLysAsnGlyGlnLysTyr 353
2747 GCGATTATATGCTGCTATTGCTACCATTTACGCCAATGGAAGATGTTT 2796
354 .....ThrLeuHisPheAspArgLe 360
2797 GGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2846
360 uGlyGlyArgMetPro...ThrTrpValIleSerProTyrSerLysLysG 376
2847 GACAAATGAAGACCGCTTACCCCTACAGTACGTTGCTTACAAACAAAGA 2896
376 IyTyrIleGluGlnTyrGlyThrAspProValThrGlyLysProAlaPro 392
2897 CGGTTATTAAAGACCATCAACACTAAT.....GGTGCAGCTTAACCCC 2937
393 Tyr.....SerAlaThrSerValLeuLysThrLeuG1 403
2938 TACCAGCCGCTTTCACACTTCAGAGCTACACACCGAGGTCAGAAAGTAAC 2987
403 yTyrLeuTrpAsp 407
2988 GCTCAAGTGGGAT 3000

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-570-311-28

seq\_documentation\_block:

; Sequence 28, Application US/08570311

; Patent No. 5824791

; GENERAL INFORMATION:

; APPLICANT: Proguiske-Fox, Ann

; APPLICANT: Tumwasorn, Somying

; APPLICANT: Lepine, Guyvaline

; APPLICANT: Han, Naiming

; APPLICANT: Lantz, Marilyn

; APPLICANT: Patti, Joseph

; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes

; NUMBER OF INVENTION: and Probes for the Detection of Periodontal Disease

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8640 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 971..6031
; US-08-570-311-28

```

alignment\_scores:

Quality: 104.50 Length: 521

Ratio: 0.488 Gaps: 30

Percent Similarity: 41.075 Percent Identity: 19.962

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x US-08-570-311-28 ..

Align seg 1/1 to: US-08-570-311-28 from: 1 to: 8640

```

17 TyrLysProGlySerLysGluSerIleGluAsnLeuLysAspLysValG1 33
1580 TACGAGCGGGCGGTACACACCGGTAGAG.....GAAACAA 1617
33 uAsnIleValTrpLeuIleLeuGluAsnArgSerPheAspAsnIleLeuG 50
1618 AAATGGTGTATGATGTCATCGTAGCCAAAAGATATGAG.....G 1658
50 IyGlyValArgArgGlnGlyLeuAspAsnProIleAsnGlyProPhe 66
1659 GAGATATATAA.....GATTTC 1675
67 CysAsnTyrLysAsnAlaSerAspProSerSerGlyLysTyrCysThrG1 83
1676 GTTGATTGGAACCAACACCGCGGT.....CTCCGTACCGA 1710
83 nAlaLysAspTyrAspSerValPheAsnAspProAspHisSerValThrG 100

```

1711 GGTGAAA.....GTGCGAAGATATTGCTCTCCCGTTACAG 1748  
|||||  
100 lYAsnAsnLeuGluPheTyr.....GlyThrTyrThrProAsnAsnGly 114  
|||||  
1749 CTAATGCTATTACAGCAGTTCGTTAGCAAGAATAACAGAAAGAAAGGTAAT 1798  
|||||  
115 AlaIleAlaSerGlyLysValAlaAspGlnSerGlyPheLeuAsnAl 131  
|||||  
1799 GATTTCGACCTATGCTTTTGGTTGGCGATCAC..... 1831  
131 aGlnLeuAsnAspTyrPro...LysLeuAlaProGluGluAlaThrArgG 147  
|||||  
1832 .....AAAGATATTCCTGCCAAAATTACTCCGGGATCAAAATCCGACC 1874  
147 InValMetGly.....TyrTyrThrGlu..... 154  
|||||  
1875 AGGTATATGGACAAATAGTAGTAAATGACCACCTACACGAAGTCTTCATC 1924  
155 .....GluGluValProThrLeuValAspLe 163  
1925 GTCGCTTCTCATGTGAGACCAAGAGGATCTGAAGACACAAATCGATCG 1974  
163 uValAspGluPhe.....ThrThrPheAsnSerTrpPheSerC 176  
|||||  
1975 GACTATTCACTAGCGCAATATAACCCAGCAAGACAAATGCTCGGT. 2023  
176 ySValProGlyProThrAsnProAsnArgLeuCys.....AlaLeu 189  
|||||  
2024 .....CAGGCTCTTTGTATGCTTCGCGCTGAA 2050  
190 AlaGlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTy 206  
|||||  
2051 GGAGCCCATCCGACAGCAATATGTTGAAAGTGATATCCAGCATGAGAATG 2100  
206 r..... 206  
2101 ANTCCGCATCTGCTACCCAGTATGGCTATACCAAGATTATCAATGTT 2150  
207 .....GlyIleSerSerLysIlePheGluAlaAlaAsnGluLys 220  
|||||  
2151 ATGATCCGGGAGTAACTCTTAAACAAATATTGATGCTTTCAAC...GGA 2197  
221 GlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAs 237  
|||||  
2198 GGAATCTCTGCTGCTCACTATACGGCCACGGTAGCGCAACACGCTTGGG 2247  
237 pSerLeuPhePhe.....ThrTyrValAsnGlnThrSerArgSerAsnV 252  
2248 TAGCTCTCACTTCGGCCACCACTCATGTGAAGCAGCTTACCAACAGCAACC 2297  
252 alValProValGluAsnPhePheGlnAspAla..... 262  
2298 AGCTACCG.....TTTATTTTCGAGCTAGCTTGTGTGAATGGCGAT 2338  
263 TyrLeuGlyValLeuProLysPhe..... 270  
2339 TTCTTATCAGCATGCGCTTCTTCGCAAGACCCCTGTATGCGTGCAAAAA 2388  
271 .....SerTyrIleAsnP 275  
2389 AGATGTAAGCGCAGGACTACTGTGCTATCATGACGCTCTACGATCAACC 2438  
275 roSer.....Cys 277  
2439 AGTCTTGGGCTTCTCTATCGCGGCGAGGATGAGATGAACGAAATTTCTG 2488  
278 CysGlyThrAsnThrAsnSerMetHisProThr...GlyAsnValSerTy 293  
|||||  
2489 TCGGAAACACCCGACACACATCAAGCGTACTTTTCGGTGTGTCAACCAT 2538  
293 rGlyGluValPheVal.....LysGlnIleTyrAspAlaIleArgG 307  
|||||

2539 GAACGGTATGTTTGTATGTTGAAAGACTATAAAAGAGGATGCTGAGAAGA 2588  
307 InGlyProGlnTrp.....AspLysThrLeuLeuPheIleThr 319  
|||||  
2589 TGCTCGACACATGACTGTTTTCGGCGACCCCTCGCTGCTGCTCGTAC 2638  
320 TyrAspGluThrGlyGlyPheTyrAspHisValProProLeuAlaVa 336  
|||||  
2639 CTT.....GTCCGACCAAAATGCAGGT 2661  
336 l...ArgProAspAsnLeuThrTyrGluThr..... 346  
2662 TAGCGCTCCGCTCAGATTAAATTCAGCGATGCTTCAGTCAACGATATCTT 2711  
347 .....AlaLysAsnGlyGlnLysTyr 353  
|||||  
2712 GCGATTATAATGGTGTATGCTACCATTTACGCCAATGGAAGATGTTTC 2761  
354 .....ThrLeuHisPheAspArgLe 360  
2762 GGTTCGTCAGTTGTCGAAATGGAACAGCTACAAATCAATCTCACAGTCT 2811  
360 uGlyGlyArgMetPro...ThrTrpValIleSerProTyrSerLysLysG 376  
|||||  
2812 GCAAAATGAAAGCAGCGCTTACCCCTTACAGTAGTTGGTTACAAACAAGAGA 2861  
376 lYTyIleGluGlnTyrGlyThrAspProValThrGlyLysProAlaPro 392  
|||||  
2862 CGGTTATTAGACCATCAACATAAT.....GGTGAGCCTAACCC 2902  
393 Tyr.....SerAlaThrSerValLeuLysThrLeuGl 403  
|||||  
2903 TACCAGCCCGTTTCCAACTTGACAGCTACAAACGCGGTCAGAAAAGTAAC 2952  
403 yTyrLeuTrpAsp 407  
2953 GCTCAAGTGGGAT 2965  
seq\_name: /cgn2\_6/ptodata/2/ins/backfiles1.seq:5223424-12  
seq\_documentation\_block:  
; Patent No. 5223424  
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD,  
; RICHARD D.  
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND  
; HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO  
; ACID SEQUENCE  
; NUMBER OF SEQUENCES: 16  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/225,032  
; FILING DATE: 27-JUL-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 78,519  
; FILING DATE: 27-JUL-1987  
; APPLICATION NUMBER: 933,107  
; FILING DATE: 20-NOV-1986  
; APPLICATION NUMBER: 902,887  
; FILING DATE: 02-SEP-1986  
; APPLICATION NUMBER: 887,140  
; FILING DATE: 17-JUL-1986  
; APPLICATION NUMBER: 823,102  
; FILING DATE: 27-JAN-1986  
; APPLICATION NUMBER: 773,430  
; FILING DATE: 06-SEP-1985  
; SEQ ID NO:12:  
; LENGTH: 1820  
5223424-12

alignment\_scores:  
Quality: 103.50 Length: 402  
Ratio: 0.588 Gaps: 22  
Percent Similarity: 43.781 Percent Identity: 21.144



## alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x 5223424-12 ..  
Align seg 1/1 to: 5223424-12 from: 1 to: 1820

```

91 PheAsnAspProAspHisSerValThrGlyAsnAsnLeu..... 103
||||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
385 TTCAATCCAGCGACTGGCAGTTAATATATCCACAACACACAGAAATAAA 434
104 .....GluPheTyr..... 106
435 CTTAGTGTAGTTTGAACAAGAAATATTCAATGTAGTACTTAAAAAATTA 484
107 .....GlyThrTyrThrPro.....AsnAsnGlyAlaIle 116
:||||| ||| :||||| :||||| :||||| :||||| :||||| :|||||
485 CAGATCAGCAACCTCCACCACCATCAAAATATATAATATGATCTAACT 534
117 AlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLe 133
||||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
535 GCACGCTAATGTCTCCACTAGACACCAATAACACACTT..... 573
133 uAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnValMetG 150
||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
574 ....CCATACACACCAGCAGCACCTAGAGTGAACA.....CTTG 610
150 lyTyrTyrThrGluGluGluValProThrLeuValAspLeuValAspGlu 166
||||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
611 GTTTTAT.....CCATGGTTACCTACA.....AAA 636
167 PheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsnPr 183
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
637 CCAACTCAATACAGATATTACCTATCATCATCAGAAACCTAAATCCACC 686
183 oAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsnA 200
| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
687 A.....ACATACACTGGACAATCACAACAAA 712
200 spAspAspPheLeuAsnTyrGlyIleSerSerLysSerIlePheGluAla 216
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
713 TAACAGACTCAATACAAACAGCAGCTACACAGTGCATATGTTCTACACA 762
217 AlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGly.....Th 230
||||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
763 ATAGAAATCCACTACCAATTCATCTTCTAAGACAGGAGATGAATTCCTC 812
230 rAsnGly.....GluPheGluProAspSerLeuPhePheThrTyrValA 245
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
813 CACAGGAATATATCACTTTGACACAAAACCAATTAATACTCACTCAT 862
245 snGlnThrSerArgSerAsnValValProValGluAsnPheGln... 260
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
863 GCACAAACAAACAGATCTCTAGAGCTGCCTCCAAACTACTAAGTGAACCT 912
261 .....AspAlaTyrLeuGlyValLeuProLysPheSerTyrIle 273
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
913 ACCACAGAGGAGACCAACACAGCAGCACTACCA..... 948
273 eaSnProSerCysCysGlyThrAsnThr.....AsnSerMetHisProThrG 289
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
949 .....GCAGCTAACACAAAGAAAGTTATCACCACCAACAA 982
289 lyAsnValSerTyrGlyGluValPheValLysGlnIleTyrAspAlaIle 305
||| ||||| ||||| :||| :||| :||| :||| :||| :||| :|||
983 TTAATAATAGCTACACAGAACA.....ACAGCACTT 1014
306 ArgGlnGlyPro...GlnTrpAspLysThrLeuLeuPheIleThrTyrAs 321
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1015 AGGCCAGCTCAGTAGGATATAATACACCATACATGAATTTTGACTACTC 1064
321 pGluThrGlyClyPheTyrAspHisValProProProLeuAlaValArgp 338
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1065 CAATGGTGGACCACTTCTA.....ACTCCTATAGTACCACACAG 1102

```

```

338 roAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThr 354
||||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
1103 CAGACACACAATATATGATGATGAACAAATGGTGTCTATAGATTATACA 1152
355 Leu.....HisPheAspArgLeuGlyGlyArgMetProTh 366
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1153 ATGGGTTACCAACATCGACACTTAACACATCTTCACAGAGCTAGAAAG 1202
366 rTrpValIleSerProTyrSerLysLysGlyTyrIle.....GluGlnT 381
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1203 ATACACATTCATCCACAAGTAATGTGAAGAGCTCCAAAGCAACAAT 1252
381 yGly..... 382
1253 TTAATCAACAGGCACCACTAAACCTAGAAAATACAAATAATGGAACACTT 1302
383 .....ThrAspProValThrGlyLysPro..... 390
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1303 TTACCTTCAGATCCAAATAGGAGGAAATCTAACACAGCATTTTCATGAATAC 1352
391 .....AlaProTyrSerAla.....ThrSerValLeuL 400
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1353 ACTCAATACATATGGACCATTAACAGCACTAAACATACTGCACCTGTAT 1402
400 yThrLeuGlyTyrLeuTrpAspIleGlu.....AspPheThrPro 413
||| :||||| ||| :||| :||| :||| :||| :||| :||| :|||
1403 TTCCAAATGGTCAATATATGGATAAAGAACTTGATACAGATCTAAACACT 1452
414 ArgVal 415
|||||
1453 AGACTA 1458

```

seq\_name: /cgn2\_5/ptodata/2/ina/5B\_COMB.seq:US-08-579-777A-1

## seq\_documentation\_block:

```

; Sequence 1, Application US/08579777A
; Patent No. 5912153
; GENERAL INFORMATION:
; APPLICANT: Enderlin, Carol S.
; APPLICANT: Selitrennikoff, Claude P.
; TITLE OF INVENTION: (1,3)B-glucan synthase genes and
; TITLE OF INVENTION: Inducible inhibition of fungal growth using the antisense
; TITLE OF INVENTION: constructs derived therefrom.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis, Graham & Stubbs L.L.P.
; STREET: 370 Seventeenth Street
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80201-0185
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,777A
; FILING DATE: 28-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,004
; FILING DATE: 18-NOV-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)892-7370
; TELEFAX: (303)893-1379
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2585 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Neurospora crassa
; STRAIN: wild-type 74-DR 23-1VA
; DEVELOPMENTAL STAGE: Mycelial
; IMMEDIATE SOURCE:
; LIBRARY: lambda zap cDNA
; CLONE: gs-1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: linkage group v
; US-08-579-777A-1

alignment_scores:
  Quality: 103.50      Length: 527
  Ratio: 0.454        Gaps: 25
  Percent Similarity: 43.264      Percent Identity: 20.683

alignment_block:
US-09-426-072-2_COPY_20_464 x US-08-579-777A-1 ..
Align seg 1/1 to: US-08-579-777A-1 from: 1 to: 2585

3 valThrSerGluThrThrSerValArgGluAlaProPheGlyThrLysPr 19
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
200 ATCGACTCTCTTACCGAAGTGGACGTCATGTCCTATTGAACCGGAACTC 249
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19 oGlySerLysGluSerIleGluAsnLeuLysAspLysValGluAsnIleV 36
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 CGGT...CTCCGGGGCTTACTACCGCTCCGACTCCGCTGCGGCGACATCA 296
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
36 alrTrpLeuIleGlu.....AsnArgSerPheAsp 46
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 ACTCGCCCTACCTCAAGGAGATGGCGGTCTTACCATGAGCTTCGAC 346
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
47 AsnIleLeuGlyGlyValArgGlnGlyLeuAspAsnProIleAsnAs 63
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
347 ACGGCTACGGTGGACGGCCATATCTCCATGCGGAGTCCCGCCCAACGG 396
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
63 nGlyProPheCysAsnTyrLysAsnAlaSerAspProSerSerGlyLys 80
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
397 TGGCCCTAT.....TCGCGCGGCTTAG 419
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 yrCysThrGluAlaLysAspTyrAspSerValPheAsnAspProAspHis 96
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
420 TGAGCCAAAGCAGTCCGTCACAGGACGATTCGAT.....GTACAC 463
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
97 SerValThrGlyAsnAsnLeuGluPheTyrGlyThrThrThrProAsnAs 113
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
464 AGCCCAACCGGC..... 475
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 nGlyAlaIleAlaSerGlyLysValAlaAspGlnSerGlyPheLeuA 130
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
476 .....GAGATCCCATGCAGA 491
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 snAlaGlnLeuAsnAspTyrProLysLeuAlaProGluGluAlaThrArg 146
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
492 ACTTCCAA...AATGGCGGTCCGCTCCCGCCACCACTGCTAGCTCCCTGG 538
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
147 GlnValMetGlyTyrThrGluGluGluValProThrLeuValAspLe 163
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
539 GAGAAATTTGACATGGCGGAGGAAATACCCCGAGTGTTCGACCA 588
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 uValAspGluPheThrThrPheAsnSer.....Trp 174
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
589 ACTTGGCGAGGCTGCACCGTCAACGATCTGACGAATCGAATACCAAGC 638
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174 heSerCysValProGlyProThrAsnProAsnArgLeuCysAlaLeuAla 190
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
639 TTGATTCG.....ACCTTCCCAAA.....GATCTCAGG 667
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

191 GlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTyrGI 207
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
668 CAATCCTTGCATAATCCAGCGCGGCGGCGGTGCTCTTCCCACTGG 717
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
207 yIleSerSerLysSerIlePheGluAlaAlaAsnGluLysGlyValSerT 224
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
718 CATATTTTTCAGCTCCATGTTCTCGACTCGAAGAGATGGTTACAGAGT 767
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 rPLeuAsnTyrAspGlyThrAsnGlyGluPheGluProAspSerLeuPhe 240
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
768 GGGAGAACTGGAAAGACGGTGAACCAAGATTATGCTGTGATCCCGTGTCT 817
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 Phe.....ThrTyrValAsnGlnThrSerArgSerAsnVa 252
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
818 GTCAAGCGACAACTCAGGCATTTCGCGCTCAGGCCTCATCAAAAGA 867
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 lValProValGluAsn.....PhePheGlnA 261
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
868 TGCCCTTAACCGCAACCAAACTGGCGACAAGAACTTCTCAACAGCAGG 917
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
261 sAlaTyrLeuGlyValLeuProLysPheSerTyrIleAsnProSerCys 277
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
918 ATTCCGTCCTCCCGCGCGGATTGAGAAAGCGTATGCGACCTCGCTGG 967
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 Cys.....GlyThrAsnThrAsnSerMetHis..... 286
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
968 ATTCTCTCTGTTGACTGGGGCGGCAACAACCTTGGCTGTGATTGGC 1017
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 .....ProThrGlyAsnValSerTyrGlyGluValPheVal...LysG 300
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1018 GCGTGGACCAAGGGCCAC.....TGGGGTCAAAATCATCTCTTTGGTC 1061
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 lIleTyrAspAla...IleArgGlnGlyPro.GlnTrpAspLysThrLeu 315
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
1062 GCGACTAGATACAAAGTACGTCGTGGCGCGCTCGTGGCGCACCTCTCT 1111
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 LeuPheIleThrTyrAspGluThrGly.....PheTyrAspHisVa 330
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1112 GCATGGTTGGCGGAGATCTCAGACGGGAGGTGGTTTGTGATGAGGA 1161
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
330 lProProLeuAlaValArgProAsnLeuThrTyrThrGluThrA 347
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1162 CACCAACAGCTCAAGCTCGCTAGTTTAAGGGCGACCGCTGTGAGCGT 1211
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 la.....LysAsnGlyGlnLysTyrThrLeuHis 356
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
1212 CTATTTTCGAGATTCTGAGGTGGAGATGATCAGAAGTATGTCGCACA 1261
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
357 PheAspArg.....LeuGlyGI 362
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
1262 GCCAACAAGCGCAAAATCTATGGCGCTTCCATGGCGTCAGCTTCGGCAT 1311
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
362 yArgMetProThrTrpValIleSerProTyrSerLys..... 374
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1312 GCGCTCTCCCCCACTCCCGGCTCTCCCTACCAAGCCCAACAGACACA 1361
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375 .....LysGlyTyrIleGluGlnTyrGlyThrAspProVal 386
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
1362 ACGAGCTCGTGGCGGTGCTACACCGCTCTCACTGGCACTTCCACCATG 1411
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 Thr.....GlyLysProAlaProTyr..... 393
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
1412 TCGATCCCATCCGACCGATTACGGAAGCAAGCCCATTCGGCGCGGT 1461
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 .....SerAlaThrS 397
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
1462 TGCTGAGGAGGACCCCCCAACCTCTCTCAGGCTAGCAACAGCTCCC 1511
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 erValLeuLysThrLeuGlyTyrLeuTrpAspIleGluAspPheThrPro 413
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||
1512 TCGAGGCCAAGCGCGGCACTTATGAGTTGATC.....ACCCA 1555
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 ArgValAlaHisSerProSerPheAspHisLeuIleGlyThrThrLeuAr 430
  :: ||| ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: |||

```





```
151 TyrThrGluGluGluValProThrLeuValAspLeuValAspGluPh 167
      ::::: ||| |||||||
700 AATCATACTGACGACTAATAAAGCATCTCTAAATGGACTTGACAAAT 749
      ::::: ||| |||||||
167 e.....ThrThrPheAsnSerTrpPheSerCysValProG 179
      ::::: ||| |||||||
750 TAAGAGTTTAGATGTAATAGCTATATATAAAGCAAAATATATATAAG 799
      ::::: ||| |||||||
179 lyProThrAsn.....ProAsnArgLeuCysAlaLeuAlaGlyThr...Ala 193
      ::::: ||| |||||||
800 GTATGACGAAATGGTTCTTCTGATCTAGTCTGTATGGCCAACTTCGAT 849
      ::::: ||| |||||||
194 AlaGlyHisGlyLysAsnAspAspPheLeuAsnTyrGlyIleSerSe 210
      ::::: ||| |||||||
850 CCAGATCATTTATCAAAAGAGTAGATAAATTTGAATTTACAAGAACA 899
      ::::: ||| |||||||
210 rLysSerIlePheGluAlaAlaAsnGluLysGlyValSerTrpLeuAsn 227
      ::::: ||| |||||||
900 TTCCTCCANTTACCACCTGTACCTAAA.....ACA 931
      ::::: ||| |||||||
227 yrAspGlyThrAsnGlyGluPheGluProAspSerLeuPhe..... 240
      ::::: ||| |||||||
932 TCGAAAATACCTCTAGCTCTATTTGACCTAGCGATCTATTTCACTATCA 981
      ::::: ||| |||||||
240 ..... 240
982 CGAGATCTCTGTAATAATTAGAATTTTCTACAAGACGGACAACGATGCT 1031
      ::::: ||| |||||||
241 .....PheThrTyrValAsnGlnThrSerArgSerAsnValValP 254
      ::::: ||| |||||||
1032 TCGAAAATTTTACTGCTGTTATCGAAACACATCTACAATCGCCTAATA 1081
      ::::: ||| |||||||
254 toValGluAsnPhePheGlnAspAlaTyrLeuGlyValLeuProLysPhe 270
      ::::: ||| |||||||
1082 CTCATGAACATACCATGTAGAT.....TTT 1107
      ::::: ||| |||||||
271 SerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMetHisPr 287
      ::::: ||| |||||||
1108 AGTTAT.....AATACCAATCT..... 1125
      ::::: ||| |||||||
287 oThrGlyAsnValSerTyrGly..... 294
      ::::: ||| |||||||
1126 .AGTGGTAATATTCAAGGCGCTCTTCAATCCGATTCGAATTCATCTTA 1174
      ::::: ||| |||||||
295 .....GluValPheValLysGlnIle 301
      ::::: ||| |||||||
1175 ATAATCCCATTTATTTCAACTCTTATAGAAATTCATTTTATAAGGCAATA 1224
      ::::: ||| |||||||
302 TyrAsp.....Al 304
      ::::: ||| |||||||
1225 CGGGATCTCTCTTTTACTTAATTTTAAAGATGGCACTCAAGGGTATGC 1274
      ::::: ||| |||||||
304 alIeArgGlnGlyProGlnTrpAspLysThrLeuLeupheIleThrTyra 321
      ::::: ||| |||||||
1275 ATTTGCCCAAGCACCA..... 1290
      ::::: ||| |||||||
321 spGluThrGlyGlyPheTyraSphis.....ValProProLeuAla 335
      ::::: ||| |||||||
1291 .....ACAGGAGTGGCTGGGACCATTCCTTTATTCGAATCTGATGGTGC 1335
      ::::: ||| |||||||
336 ValArgProAspAsnLeuThrThrGluThrAlaLysAsnGly..... 350
      ::::: ||| |||||||
1336 CCAGAAGGCGATAAATAAACTATATTTATCTCTCCAGGTGATACATT 1385
      ::::: ||| |||||||
351 .....GlnLysTyrThrLeuHisPheAspArgLeuGlyAla 363
      ::::: ||| |||||||
1386 AAGATTTTCATCATGTATATCTCTTATAGTACTCCAACATAAATG 1435
      ::::: ||| |||||||
363 rgMetProThrTrpValIleSerProTyr...SerLysLysGlyTyrIle 378
      ::::: ||| |||||||
1436 AACTATCAACAGAAAAATCAAGGCTTCTCGCAAAAGGATATATC 1485
      ::::: ||| |||||||
379 GluGlnTyrGlyThrAspProValThrGlyLysPro 390
```

```
1486 AAAAAATCAAGGGATCATGAAATATATTACGGTAAACCA 1521
      ::::: ||| |||||||
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2001, 16:21:42 ; Search time 53.66 Seconds  
(without alignments)  
310.020 Million cell updates/sec

Title: US-09-426-072-2\_COPY\_20\_264

Perfect score: 1323

Sequence: 1 SPVTSEYTVREAPFGYKPG.....NYDGTNGEFPDLSLFTTYVN 245

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	210.5	15.9	514	2 T02648	hypothetical prote
2	174.5	13.2	462	2 T46205	hypothetical prote
3	162.5	12.3	730	2 A26391	phospholipase C (E
4	162.5	12.3	730	2 B83540	hemolytic phosphol
5	161.5	12.2	508	2 F70662	probable plcC prot
6	148	11.2	512	2 G70662	probable plcB prot
7	145.5	11.0	512	2 H70662	probable plcA prot
8	133	10.1	692	2 A36143	phospholipase C (E
9	133	10.1	692	2 E83230	non-hemolytic phos
10	128	9.7	221	2 T35551	probable phospholi
11	96	7.3	422	2 B31776	hypothetical prote
12	95.5	7.2	2178	2 S55805	alpha-toxin - Clos
13	93	7.0	867	2 A49284	immediate-early pr
14	92.5	7.0	428	2 S32228	glutamate--ammonia
15	92.5	7.0	1288	2 E71884	vacuolating cytoto
16	92	7.0	737	2 D40506	PRF4 protein - hum
17	90.5	6.8	1146	2 S07915	RF2 protein - yeas
18	90.5	6.8	1472	2 H82802	fimbrial assembly
19	90	6.8	600	2 S30153	low-temperature-in
20	90	6.8	760	2 S73756	probable lipoprote
21	89	6.7	704	2 H71523	hypothetical prote
22	88.5	6.7	315	1 AJYLN8	glutamate--ammonia
23	88.5	6.7	1217	2 S52714	sericin1B - silkw
24	87.5	6.6	461	2 S42084	basic protein - Cl
25	87.5	6.6	960	1 A36551	discs-large tumor
26	87.5	6.6	3194	2 D71917	toxin-like outer m
27	86.5	6.5	397	2 E69858	conserved hypothet
28	86.5	6.5	480	2 T24087	hypothetical prote
29	86.5	6.5	1211	2 S54500	alpha, alpha-trehal

30 85.5 6.5 434 1 AJBHQ glutamate--ammonia  
31 84.5 6.4 325 2 E64998 hypothetical prote  
32 84.5 6.4 356 2 S71580 glutamate--ammonia  
33 84.5 6.4 557 2 A57050 K-glypican precurs  
34 84.5 6.4 3005 2 S33642 homeotic protein z  
35 84 6.3 1290 2 A57190 eberlein precursor  
36 83.5 6.3 192 1 UCFRPP cuticle protein pr  
37 83.5 6.3 427 2 S42658 H+-transporting At  
38 83.5 6.3 476 2 S09489 carboxypeptidase H  
39 83.5 6.3 533 2 A37180 chromogranin/secre  
40 83.5 6.3 624 2 T02289 probable polygalac  
41 83.5 6.3 820 2 B72575 hypothetical prote  
42 83.5 6.3 2893 2 A64556 toxin-like outer m  
43 83 6.3 474 2 H71126 probable 4-aminobu  
44 83 6.3 488 2 A32235 poly(3-hydroxybuty  
45 83 6.3 1026 1 A40315 maternal effect pr

## ALIGNMENTS

RESULT 1

T02648 hypothetical protein F12C20.9 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 14-May-1999

C:Accession: T02648

R:Rounsley, S.D.; Ronning, C.M.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.;

submitted to the EMBL Data Library, August 1998

A:Description: Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.

A:Reference number: Z14685

A:Accession: T02648

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-514 <ROU>

A:Cross-references: EMBL:AC005168; NID:g3426033; PID:g3426039

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Introns: 156/2; 280/3

A>Note: F12C20.9

Query Match 15.9%; Score 210.5; DB 2; Length 514;  
Best Local Similarity 30.3%; Pred. No. 1.3e-09;  
Matches 61; Conservative 37; Mismatches 76; Indels 27; Gaps 6;

QY 32 VENIVWLILNRSFONILGGVRRGGLDNPIN--NGPFCNYKNASDPSSGKYCTQAKDVS 89

Db 27 IKTVVYVWENRSFDMHLMGWMKK--LNPENGVDGSESNPVSDPSRRKIFGSGSH-Y 83

QY 90 VFNDPHSHVYVGNLLEFYCY-----TPNNGAIAAGKVAVDQSGFLNAQLNDYPKLAPEE 143

Db 84 VDDPDGHSFAIQREQVFGSNDTSMDDPPPMNGFVQQA-YSEDPGNGMSA----- 130

QY 144 ATROVMGYTTEERPTLDVDEFTFNFSWFCVGPPTNPRLCALAGTAGAGKNDODF 203

Db 131 ---SVMMNGFPPDKVPYKSLVSEFAVDFRWFASVPSSTQPNRMFVHSGTSAGATNNPIS 187

QY 204 LNYGISKSIIFEAANEKGVSW 224

Db 188 LAKGYPORTIFDNLDEEFSF 208

RESULT 2

T46205 hypothetical protein T8P19.120 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000

C:Accession: T46205

R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;

submitted to the Protein Sequence Database, December 1999

A:Reference number: Z23008





Query Match 12.2%; Score 161.5; DB 2; Length 508;  
Best Local Similarity 27.3%; Pred. No. 1.3e-05;  
Matches 67; Conservative 27; Mismatches 92; Indels 59; Gaps 11;  
Qy 10 VRAPGKPGKESIEENLKVNIWVLIENSRFDNLGGVRR-OGLDNPNNGPFCN 68  
Db 26 VIEKAYGAGPCS---GHLTD-IEHIVLCLOENSRFDHVGTLSDVDFDPT---PLFQ 77  
Qy 69 YKNASDPSSGKCYCTQAKDYDV-----FNDPDHVSVTGNLFYCYTYT 110  
Db 78 QKQWPE-----TQALDPGTLPRINTTGGPGVGCVDNDPDHQAHLW----- 126  
Qy 111 PNNATASGKVVADQSGFLNAQNDYPKLAPAEATROVMGYTEEEVPTLVLDVDEFTTF 170  
Db 127 -NGGA-----NDGWLPAQ-----ARTRSVANTPVVMGYTARPDIPHYLLADTFITC 172  
Qy 171 NSWFSCVCPNPNRLCALAGTAAGH-KNDDDFLNYGTSKSF-----EANEKGV 223  
Db 173 DQYFSSLLGGTMPNRLWISATVNPDDGGGQGVPAIVPAOPKLTFTWRIMPONLSDGIS 232  
Qy 224 WLNYD 228  
Db 233 WKVYN 237  
RESULT 6  
G70662  
probable plcB protein - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: G70662  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A: Reference number: A70500; MUID: 98295987  
A: Accession: G70662  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-512 <COL>  
A: Cross-references: GB:283860; GB:ALL23456; NID:g3261681; PIDN:CAB06147.1; PID:g1781257  
A: Experimental source: strain H37Rv  
C: Genetics:  
A: Gene: plcB

Query Match 11.2%; Score 148; DB 2; Length 512;  
Best Local Similarity 25.9%; Pred. No. 0.00017;  
Matches 58; Conservative 25; Mismatches 57; Indels 84; Gaps 10;  
Qy 32 VENIWLILENRSFDNLGGVR-ROGLD-----NPNNGPFCNKASDPSS 77  
Db 43 IEHIVLLMOENSRFDHVGTLSTRGFDFTTPPVFAQSGWNPMT-----QAVDPAG 94  
Qy 78 GKCYCTQAKDYDV-----FNDPDHVSVTGNLFYCYTYTPNNGAIAAGKVVADQSGF 128  
Db 95 ---VTLPYRFDTRGPLVAGECVNDPDHSGWGMNSW-----NGGA-----NDNW 136  
Qy 129 LNAQLNDYPKLAPAEATROVMGYTEEEVPTLVLDVDEFTTFNSWFCVPGTNPRLCA 188  
Db 137 LPAQGFSPLOGNVPVT---MGFTYRDLPLTHYLLADTFVCDVGCFLSGTTPNRL-- 191  
Qy 189 LAGTAAGHGKNDODFLNYGTSKSFEEANEKGVSWLNYDGTNG 232  
Db 192 -----YWMS-----AWIDPDGTDG 205

RESULT 7  
H70662  
probable plcA protein - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70662  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A: Reference number: A70500; MUID: 98295987  
A: Accession: H70662  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-512 <COL>  
A: Cross-references: GB:283860; GB:ALL23456; NID:g3261681; PIDN:CAB06148.1; PID:g1781257  
A: Experimental source: strain H37Rv  
C: Genetics:  
A: Gene: plcA  
Query Match 11.0%; Score 145.5; DB 2; Length 512;  
Best Local Similarity 27.5%; Pred. No. 0.00027;  
Matches 66; Conservative 21; Mismatches 78; Indels 75; Gaps 10;  
Qy 32 VENIWLILENRSFDNLGGVRROGLDNPINNPGFCNKASDPSSGKY---CTQAKDY 87  
Db 43 IEHIVLLMOENSRFDHVGTLSS-----TNGFNAASPAFQMGWNPWTQALDP 90  
Qy 88 DSV-----FNDPDHVSVTGNLFYCYTYTPNNGAIAAGKVVADQSGFL 129  
Db 91 AGVTIPFRLDTRGPFGLDGEVNDPDHSGWGMHLAW-----NGGA----- 130  
Qy 130 NAQLNDYPKLAPAEATROQ-----VMGYTEEEVPTLVLDVDEFTTFNSWFCVPGPTN 182  
Db 131 ---ND-NWLPQAATTRAGPVVPLTMGYTYRQDIPHYLLADTFITCDGVCFLSGTTL 184  
Qy 183 PNRICAL-----AGTAAGHGKNDODFLNYGTSKSF-EANEKGVSWLNYDGTN-GEF 234  
Db 185 PNRLYWLSANIDPAGTDGGPQLVEFGFLPQOFSWRIMPENLEDAGVSKVYQNKGLGRF 244  
RESULT 8  
A36143  
phospholipase C (EC 3.1.4.3), nonhemolytic - Pseudomonas aeruginosa  
C:Species: Pseudomonas aeruginosa  
C:Date: 30-Nov-1990 #sequence\_revision 30-Nov-1990 #text\_change 08-Oct-1999  
C:Accession: A36143  
R: Ostroff, R.M.; Vasil, A.I.; Vasil, M.L.  
J. Bacteriol. 172, 5915-5923, 1990  
A: Title: Molecular comparison of a nonhemolytic and a hemolytic phospholipase C from  
A: Reference number: A36143; MUID: 91008968  
A: Accession: A36143  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-692 <OST>  
A: Cross-references: ENBL:M59304; NID:g151495; PIDN:AAA25968.1; PID:g151496  
C: Keywords: phosphoric diester hydrolase  
Query Match 10.1%; Score 133; DB 2; Length 692;  
Best Local Similarity 25.8%; Pred. No. 0.0042;  
Matches 60; Conservative 32; Mismatches 89; Indels 52; Gaps 11;  
Qy 27 NLKDKVENTVWLIENSRFDNLGG---GVRQK---LDNPNNGPFCNKASDPSSGKY- 80  
Db 44 NLKD-VEHVIVILMOENSRFDHVGTLKGVGRGDMRMAIPLPDQGRVWQKSGKEILPYH 102  
Qy 81 ---CTQAKDYDSVFNPDHVSVTGNLFYCYTYTPNNGAIAAGKVVADQSGFLNAQLNDY 136  
Db 103 FDTSTSAQRVDG-----TPHTWPD-----QOAWNGRMD-- 133  
Qy 137 PKLAPAEATROVMGYTEEEVPTLVLDVDEFTTFNSWFCVPGTNPRLCALAGT---A 193  
Db 134 -KWLPAK-TERSLGYKEQDIAQFAMANAFTICDAYHCSFGGTGTPNRLFLWTGTNDPL 191

14 PFGYKPGSKESIENLKDKNIVWLILENRSFDNIIIG--GVRROGLDNPINNGPFCNYK 70

A; Molecule type: DNA



A:Gene: vaca

```
Query Match      7.0%; Score 92.5; DB 2; Length 1288;
Best Local Similarity 22.0%; Pred. No. 20;
Matches 63; Conservative 37; Mismatches 95; Indels 91; Gaps 14;

Qy  4 TSEYTSVREAPFGYKP-----GSKESIENLKDKNENIWL-----ILE 41
Db  851 SANYALVKNAFPAHSATPNLVAIQHDFGTIESVFELANRSKDIDTLTHTSGAQQGRDLQ 910
Qy  42 NRSFDNILLGVRQGLDNPINNGPCNVKNASDPSSGKYCTQAKDYDSVFNDP---DHSV 98
Db  911 TLLIDSHDAGYARQMDN-----TSTGEITKQLNAATDALNNVASLEHKQ 955
Qy  99 TGNNEFYGYTTPNNGAIASGVVADSGFLNAQLNDY-----PKLAPEEATQVM 149
Db  956 SG-----LQTLISLSNAMILNSRLVLSRKHTN--HINFAQLQALKGQEFASLESAAEVL 1009
Qy  150 GYITEE-EVPTLVLDVDEFTTFNSFCVPGPT-NPNRLCALAGTAAGHGKNDLDFLN-- 205
Db  1010 YQFAPKYEKPT-----NWNANAIGGASLNSGNSASLYGTSAG-----VDAFLNGN 1054
Qy  206 -----YGISSKSIFEAANEKGVSWLNDGTNGEPEFEDSLFF 241
Db  1055 VEAIVGGFGSGYGSFES--NQANS-----LNSGANNANFGVYSRFF 1093
```

Search completed: February 26, 2001, 16:21:44  
Job time: 41384 sec





QY 32 VENIWLILNRSFONILGGYR--ROGLNPNNGPCNYKNASDPSSCKYCTQAKDYDS 89  
 Db 14 IKTIVLVQENRSFDTLGGFKELNREIDGVTKSDPKNSVSSDTSNLSR--VFGDQSQ 71  
 QY 90 VFN-DPDHSTGNLEFY-----GTYTNPNGAIASGKVVADQSGFL-NAOLNDYPKLAP 141  
 Db 72 YVNDPCHSITDIYEQVFGKPDGSGKPDNPGH-----PNMSGFAQNAERN-----K 118  
 QY 142 EEAATVOMGYTTEEEVPTLVLDVDEFTTNSWFCVPGPTNPRLCALACTAAGHKND 201  
 Db 119 KGMSSAVMNGFNPALPVYKELVQNFACIDRWFASVPASTQPNRLVHSGATSHGATSNK 178  
 QY 202 DFLNYGSSKSIFFAANEKGVSWLNYDGTNGEFPDLSFF 241  
 Db 179 KLLGEGFPQKTFISLDEAGFSF-----GIYYQFPSTLFF 214

## RESULT 5

Q9S816  
 ID Q9S816 PRELIMINARY; PRT; 521 AA.  
 AC Q9S816;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE PUTATIVE PHOSPHOLIPASE.  
 GN T12J13.18 OR T21P5.4.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN 11  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=CV. COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,  
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.B.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RT "Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC009327; AAF03477.1;  
 DR EMBL; AC009895; AAF01581.1;  
 SQ SEQUENCE 521 AA; 59071 MW; 1A39ACEF6807B8EF CRC64;

Query Match 14.4%; Score 190.5; DB 10; Length 521;  
 Best Local Similarity 29.1%; Pred. No. 3.4e-08;  
 Matches 67; Conservative 31; Mismatches 83; Indels 49; Gaps 10;  
 QY 32 VENIWLILNRSFONILG-----GVRQ-----GLONPIN--NGPFCYKNAS 73  
 Db 13 IKTIVLVQENRSFDTLGGFKELNREIDGVTKSDPKNSVSSDTSNLSR--VFGDQSQ 72  
 QY 74 DPSSGKCTQAKD-YDSVNDPDHSTGNLEFYCTYTPNNGAIASGKVVADQSGFL-NA 131  
 Db 73 DPNPGH---SIRDIYEQVFGKPDMS-----GHPDPNPG-----ATMSGFAQNA 113  
 QY 132 QLDYKPLAPEATRVQVMGYTTEEEVPTLVLDVDEFTTNSWFCVPGPTNPRLCALAG 191  
 Db 114 ERK-----MKGSSAVMNGFKPDALPVYKELVQNFACIDRWFASVPASTQPNRL 167  
 QY 192 TAAGHKNDLFLNGISSKSIFFAANEKGVSWLNYDGTNGEFPDLSFF 241  
 Db 168 TSHGTTNNKRLILIEGFPQKTFISLDEAGFTF-----GIYYQFPSTLFF 213

## RESULT 6

Q9XB13  
 ID Q9XB13 PRELIMINARY; PRT; 514 AA.  
 AC Q9XB13;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DE  
 GN  
 OS  
 OC  
 OX  
 RN  
 RC  
 RP  
 RA  
 RT  
 RL  
 DR  
 SQ

DE PHOSPHOLIPASE.  
 GN RVDZ-PLCD.  
 OS Mycobacterium bovis BCG.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=33892;  
 RN 11  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=BCG-PASTEUR;  
 RX MEDLINE=99255698; PubMed=10320585;  
 RA Gordon S.V., Brosch R., Billault A., Garnier T., Eiglmeier K.,  
 RA Cole S.T.;  
 RT "Identification of variable regions in the genomes of tubercle bacilli  
 RT using bacterial artificial chromosome arrays."  
 RL Mol. Microbiol. 32:643-655(1999).  
 DR EMBL; Y18606; CAB4656.1;  
 SQ SEQUENCE 514 AA; 55770 MW; 9D99A466D312BA30 CRC64;

Query Match 13.5%; Score 178.5; DB 2; Length 514;  
 Best Local Similarity 28.5%; Pred. No. 3.5e-07;  
 Matches 75; Conservative 25; Mismatches 96; Indels 67; Gaps 11;  
 QY 10 VREAPFGYKPGKESINLKDVENIWLILNRSFONILGGVRRGGLDNPINNGPCNY 69  
 Db 33 VIEKAYGAGPCS-----GHLTD-IEHFVFPQENRSFDHYFGL--SGTGGFTVSPFLQ 85  
 QY 70 KNASDPSSCKYCTQAKDYDSV-----FNDDHSTGNLEFYCTYTP 111  
 Db 86 KGNWP-----MTQALDATGVTMPYRDTTRGPFLDGACVNDPDHS-----WVAMHES 132  
 QY 112 NNGAIASGKVVADQSGFLNADNDYKPLAPEEATRVQVMGYTTEEEVPTLVLDVDEFTT 171  
 Db 133 WNGGV-----NDNWLPAQ-----AKTRSAATPTVMGYTTRDIPHIYLAFTVCD 180  
 QY 172 SWFCVPGPTNPRLCALAGTAAGHKNDLFL-----NYGISKSIFFAANEK 220  
 Db 181 RFYCSVLGPTNRLNRLWLSATIDPDGNGGPELQSPFTQPVRRFGWRIMPQLSDA---- 236  
 QY 221 GVSMLNYDGTNGEFPDLSFFTY 243  
 Db 237 GVSWKYV--RNKTLGPISSVLT 257

## RESULT 7

Q9SMN8  
 ID Q9SMN8 PRELIMINARY; PRT; 462 AA.  
 AC Q9SMN8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE HYPOTHETICAL 51.3 KDA PROTEIN.  
 GN T8P19.120.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN 11  
 RC SEQUENCE FROM N.A.  
 RP Choise N., Robert C., Brottier P., Wincker P., Cattolico L.,  
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN 12  
 RC SEQUENCE FROM N.A.  
 RP EU Arabidopsis sequencing project;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL133315; CAB62350.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 462 AA; 51284 MW; BE6492B529991095 CRC64;

```
Query Match      13.2%; Score 174.5; DB 10; Length 462;
Best Local Similarity 29.0%; Pred. No. 6.7e-07;
Matches 54; Conservative 25; Mismatches 82; Indels 25; Gaps 6;

QY 59 NPINN---GPCNKNASDPSSGKYCTQAKDYDSVFNPDHSHVGTGNLNFYGYTFNNGCA 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 NPTINGVTGQECN----PVPNSTQITCFTDAEFVDPDGHSGFAVEQQVFGS---GPGQ 58

QY 116 IASGKVADQSGFLNALQNDYPKLAPEATROVMGYTYTEEEVPTLVDLVDDEFTNSWFS 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 IPS-----MMGFEQALS-----MPQNLSETVMKGFPEAPVVAELVKEFAVDFRWS 107

QY 176 CVPGPTNRLCALAGHCKNDLFLNYGSSKSFIFAEANKEGVSRLNYDINGEET 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 SIPTQPNRFLVYSATSHGSTSHVKQLAQGPQKTIPTFDSLSHNDIDF----GIYFQNI 163

QY 236 PDSLEF 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 PTTLEY 169

RESULT 8
P95245 PRELIMINARY: PRT; 508 AA.
AC P95245;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PLCC.
GN PLCC OR RV2349C OR MTCY98.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter K., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; Z83860; CAB06146.1; -.
DR TUBERCULIST; RV2349c; -.
DR NCBI_TaxID=1773;
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter K., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; Z83860; CAB06146.1; -.
DR TUBERCULIST; RV2349c; -.
DR NCBI_TaxID=1773;
RP SEQUENCE 508 AA; 55047 MW; A0A517ABD20BBA2 CRC64;

Query Match      12.2%; Score 161.5; DB 2; Length 508;
Best Local Similarity 27.3%; Pred. No. 9.8e-06;
Matches 67; Conservative 27; Mismatches 92; Indels 59; Gaps 11;

QY 10 VREAPFGYKSGKESIKKVENIVLILNRSFDNLGLGVRR-QGLDNPINNGPFCN 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26 VIEKAYAGAPCS---GHLTD-IEHIVLCIQENRSPDHYFGTLTSAVDGFTPT---PLFQ 77

QY 69 YKNASDPSSGKYCTQAKDYDSV-----FNDPDHSHVGTGNLNFYGYTYT 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 QKWNPE-----TQALDPTGITLPYRINTGPGNCGVGVNDPDHQTAAHLSW----- 126

QY 111 PNNGAIASGKVADQSGFLNALQNDYPKLAPEATROVMGYTYTEEEVPTLVDLVDDEFTTF 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 -NGGA-----NDGWLPAQ---ARTRSVANTPVVMGYARPDIPHYLLADTFETIC 172

QY 171 NSWFSCVPGPTNRLCALAGTAGHCKNDLFLNYGSSKSFIFAEANKEGVS 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 DOYFSSLLGGTMFNRLYISATVNPDGQGGPQVPEIPALQPKLTFTWRIMPQLNSDAGIS 232
```

```
QY 224 WLNVD 228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 MKVYN 237

RESULT 9
P95246 PRELIMINARY: PRT; 512 AA.
AC P95246;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PLCC.
GN PLCC OR RV2350C OR MTCY98.19C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter K., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; Z83860; CAB06147.1; -.
DR TUBERCULIST; RV2350c; -.
DR NCBI_TaxID=1773;
RP SEQUENCE 512 AA; 55623 MW; B82672FA0C81DA9B CRC64;

Query Match      11.2%; Score 148; DB 2; Length 512;
Best Local Similarity 25.9%; Pred. No. 0.00014;
Matches 58; Conservative 25; Mismatches 57; Indels 84; Gaps 10;

QY 32 VENIVLILNRSFDNLGLGVRR-ROGLD-----FNDPDHSHVGTGNLNFYGYTYTPNNGAIASGKVADQSGF 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43 IEHIVLILNRSFDNLGLGVRR-ROGLD-----FNDPDHSHVGTGNLNFYGYTYTPNNGAIASGKVADQSGF 94

QY 78 SKYCTQAKDYDSV-----FNDPDHSHVGTGNLNFYGYTYTPNNGAIASGKVADQSGF 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 ---VTLFYREDTTRGFLVAGECVNDPDHSHVGTGNLNFYGYTYTPNNGAIASGKVADQSGF 136

QY 129 LNAQLNDYPKLAPEATROVMGYTYTEEEVPTLVDLVDDEFTNSWFSVPGPTNRLCA 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 LPAQVPSPLQGNVPT---MGFYTRDLPIHYLLADTFVCDGFCYFGLLGGTTNRL-- 191

QY 189 LAGTAAGHCKNDLFLNYGSSKSFIFAEANKEGVSRLNYDINGTNG 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 -----YWMS-----AWIDPDGTDG 205

RESULT 10
Q50561 PRELIMINARY: PRT; 521 AA.
AC Q50561;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE PHOSPHOLIPASE C.
GN MPCB.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
```







```
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN=ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; 248636; CAA88565.1; -
DR EMBL; 223280; CAA80818.1; -
DR EMBL; 223281; CAA80819.1; -
DR INTERPRO; IPR002479; -
DR PFAM; PF01473; CW_binding_1; 10.
SQ SEQUENCE 2178 AA; 250134 MW; 9B0ADCE031C4A75A CRC64;

Query Match          7.2%; Score 95.5; DB 2; Length 2178;
Best Local Similarity 23.5%; Pred. No. 30;
Matches 70; Conservative 33; Mismatches 104; Indels 91; Gaps 16;

Qy 3 VTSEYTSVREAPFGYKPGSKESIE-----NLKQKVEN---IVWLILEN----- 42
Db 1378 INNIFSTLKVNDGFKIG-KQFISIKNTPRAINLSFKINNIVIVSYLNHEKSNSTII 1436
Qy 43 -----RSTDNILGGVRRQGL----DNIPNPGPCNKYKNASDPSSGKYCTQAKD---- 86
Db 1437 SSDLNDIKNNFDNLLDNINYIGLSISDNTIN----CTVRNDEVYMECKIFLNEKLVFI 1492
Qy 87 -----YDSVFNDPDHVSVTG--NNLEFYGYTPNNGAIGSKVYVADQSGFLNAQLNDY 136
Db 1493 QNELEHLHYDSVNSKDSQYLINNPINNYYKY-----KGYIVEGTFL-----INSTENKY 1541
Qy 137 PKLAPEATROVMGYYTEEEV-----PTLVLDVDEFTFNSW---FSCVP 178
Db 1542 -SLYIENKIMLKGYLESSVFKTIODKIYSKEKVNDYILSLIKKFTVNIQLCPFMIVS 1600
Qy 179 GPTNPNR-----LCALAGTAAGHKNDDDLNYGISKSIFFAAN--EKGYSWLNYDG 229
Db 1601 GVDENNRYLEYMLSTNNKWIINGGYWENDFNKYIVD---FEKCNVIVSGSNKUNSEG 1655
```

Search completed: February 26, 2001, 19:32:39  
Job time: 11520 sec





Db 95 NYKYEF-PYHWDT-----KVTSQWVSSQNHWSAFHATWQGRDKWM 138  
 QY 136 ---YKLAPEATQVMGYTTEEEVPTLVLDVDEFTTNSWFSVCPGTPNRLCALAGT 192  
 Db 139 AVQYF-----EANGYKRGDIPYYALADAFILCEAYHQSMGTPNRLHMSGR 189  
 QY 193 AAGCHKNDDELNGYTKSKSIF-----EANEKGVSWLNYDGTNGEFPDPSLF 241  
 Db 190 AAPSGDKRKHGNDMGDTGASCTVDWTTYPRLSAGVDMRVYQ--EGGYRSSLWY 247  
 QY 242 TYVN 245  
 Db 248 LYVD 251  
 RESULT 2  
 ID PHLN\_PSEAE STANDARD; PRT; 692 AA.  
 AC P15713;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-MAY-1991 (Rel. 18, Last annotation update)  
 DE NON-HEMOLYTIC PHOSPHOLIPASE C PRECURSOR (EC 3.1.4.3)  
 GN (PHOSPHATIDYLCHOLINE CHOLINEPHOSPHOHYDROLASE).  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 RN Pseudomonas.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=91008968; PubMed=2120196;  
 CC Ostroff R.M., Vasili A.I., Vasili M.L.;  
 CC "Molecular comparison of a nonhemolytic and a hemolytic phospholipase  
 CC C from Pseudomonas aeruginosa."  
 CC J. Bacteriol. 172:5915-5923(1990).  
 CC !- FUNCTION: HYDROLYZE PHOSPHATIDYL SERINE AS WELL AS  
 CC PHOSPHATIDYLCHOLINE.  
 CC !- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = 1,2-  
 CC DIACYLGLYCEROL + CHOLINE PHOSPHATE.  
 CC !- SIMILARITY: 40% IDENTITY TO HEMOLYTIC PHOSPHOLIPASE C.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M59304; AAA25968.1; -  
 CC EMBL; M30223; AAA25969.1; -  
 CC PIR; A36143; A36143.  
 CC Hydrolyase; Signal.  
 FT SIGNAL 1 35  
 FT CHAIN 36 692 NON-HEMOLYTIC PHOSPHOLIPASE C.  
 SQ SEQUENCE 692 AA; 77019 MW; E3A2AD593DE0B12B CRC64;

Query Match 10.1%; Score 133; DB 1; Length 692;  
 Best Local Similarity 25.8%; Pred. No. 0.001;  
 Matches 60; Conservative 32; Mismatches 89; Indels 52; Gaps 11;  
 QY 27 NLKDKVENIWLILNRSFDNLTG--GVRROG--LDNPNNGPFCNYKNASDPSSGKY- 80  
 Db 44 NLKD-VEHVILMQENRSFDHYFTGLKGVNGFGDRAIPLPDGQVWVHQKSGKGLLPYH 102  
 QY 81 ---CTQAKDYDSVFNPDHVSVTGNLLEFYGYTPNNGATASGVVADQSGFLNAQLNDY 136  
 Db 103 EDTSTTSQVRDVG-----TPTWEDA-----QQAWEGRMD-- 133  
 QY 137 PKLAPEATQVMGYTTEEEVPTLVLDVDEFTTNSWFSVCPGTPNRLCALAGT---A 193  
 Db 134 KWLPAK-TERSLGIYKEQDIAFOFAMANAFTICDACHCSFQSGTGNPRLFLWTGNDPL 191

QY 194 AGHG----KNDDF---LNVGISKSTFEAANEKGVSWLNYDGTNGEFPDSJ 239  
 Db 192 GQHGPGPTNHDHSGNPGVEQGYTWTYPERLQAAGITWRVYQDMAONFSDNPL 244  
 RESULT 3  
 ID PHOX\_KLULA STANDARD; PRT; 421 AA.  
 AC P08540;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE POTENTIAL ACID PHOSPHATASE (EC 3.1.3.2).  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Kluyveromyces.  
 RN [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;  
 CC MEDLINE=89034156; PubMed=3053697;  
 CC Chang Y.-D., Dickson R.C.;  
 CC "Primary structure of the lactose permease gene from the yeast  
 CC Kluyveromyces fragilis. Presence of an unusual transcript structure."  
 CC J. Biol. Chem. 263:16696-16703(1988).  
 CC !- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN  
 CC ALCOHOL + ORTHOPHOSPHATE.  
 CC !- SIMILARITY: VERY STRONG, TO ASPERGILLUS NIGER AND PENICILLIUM  
 CC CHRYSOGENUM PHOA.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X06997; CAA30054.1; ALT\_INIT.  
 CC PIR; B31776; B31776.  
 CC Hypothetical protein; Hydrolase.  
 FT ACT\_SITE 228 228 PROTON DONOR (BY SIMILARITY).  
 SQ SEQUENCE 421 AA; 47193 MW; 2F7E614C943824F6 CRC64;  
 Query Match 7.3%; Score 96; DB 1; Length 421;  
 Best Local Similarity 23.7%; Pred. No. 0.73;  
 Matches 33; Conservative 24; Mismatches 50; Indels 32; Gaps 6;  
 QY 103 LEFYGYTTPNNGAIAS-----GKVADQSG-----FLNAQLNDYPKLAPEATROV 148  
 Db 31 LRTYSTISPSLSIEISASASATEVAEVSDVEGAFAKFFIIFLENTDYDKAGDES---- 86  
 QY 149 MGYTTEEVPTLVLDVDEFTTNSWFSVCPGTPNRLCALAGTAAGCHKNDDELNGY 208  
 Db 87 LSWLAQGI-----TLTNW--ALHTPSEPNYLSVGGDY--FALDDDRFISMS 132  
 QY 209 SSKSIFEAAANEKGVSWLNY 227  
 Db 133 NVSNIVDLLDTKGISWAEY 151  
 RESULT 4  
 ID GPC4\_HUMAN  
 AC O75487; Q9UPD9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GLYPICAN-4 PRECURSOR (K-GLYPICAN).  
 GN GPC4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE=99005528; PubMed=9787072;
RX Veugleers M., Vermeesch J., Watanabe K., Yamaguchi Y., Marynen P.,
RT David G.;
RA "GPC4, the gene for human K-glypican, flanks GPC3 on Xq26: deletion of
RT the GPC3-GPC4 gene cluster in one family with Simpson-Golabi-Behmel
RT syndrome.";
RL Genomics 53:1-11(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Pilia G., Mazzarella R., Huber R., Crisponi L., Lindsay S.,
RA Ireland M., Cao A., Schlessinger D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN. MAY BE
CC INVOLVED IN THE DEVELOPMENT OF KIDNEY TUBULES AND OF THE CENTRAL
CC NERVOUS SYSTEM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF064826; AAC69991.1; -.
DR EMBL: AF064826; AAC31899.1; -.
DR MW: 300168; -.
DR INTERPRO: IPR001863; -.
DR PFAM: PF01153; Glypican; 1.
DR PROSITE: PS01207; Glypican; 1.
KW Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.
FT SIGNAL 1 18
FT CHAIN 19 529
FT PROPEP 530 556
FT LIPID 529 529
FT CARBOHYD 514 514
FT CARBOHYD 494 494
FT CARBOHYD 498 498
FT CARBOHYD 500 500
FT CONFLICT 442 442
FT SEQUENCE 556 AA; 62426 MW; CD4B1CDE7A95D80D CRC64;
QY 47 NILGG-VRRQG-LDNPINN-----GPFCKYKNASDPSSGKYCTQAKDYDSVF 91
DB 272 NIMRGCLANGLDGLDPEWNTIDAMLVABERLEGPP-NIESVMDP-----IDVKISDAIM 324
QY 92 NDPDHSVTCNNLEFFGYTPNNGAIAAGV--VADQSCFLNAQINDPKLAPEATQVM 149
DB 325 NMQDNSVQVSQKVFQGGCPK--PLPAGRISRSISSEAF-SARFRPH---HPEERPTAA 378
QY 150 GYETEVEVTLVDLDEPTFNWFSVCPGPTNPNRLCALAGTAAGKNDGDDFLNYGIS 209
DB 379 GTSILRLVTDVKDLKQAKFF--WSSL-----PSNVCNDRMAAGNG-NEDDCWNGKGK 429
QY 210 SKSIFFAANEKGVSWLNDGTNGEPEFDS 238
DB 430 SRYLFAVTGNGLVN-----OGNNPEVQVDT 454
RESULT 5
KTXA_KLUJA

```

```

ID KTXA_KLUJA STANDARD; PRT: 1146 AA.
AC P09805;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE KILLER TOXIN ALPHA/BETA SUBUNIT PRECURSOR (RF2 PROTEIN) [CONTAINS:
DE KILLER TOXIN ALPHA SUBUNIT; KILLER TOXIN BETA SUBUNIT (ENDOCHITINASE)
DE (EC 3.2.1.14)].
OS Kluyveromyces lactis (Yeast).
OG Plasmid pGKL-1.
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Kluyveromyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 2359/152;
RA Sor F., Fukuhara H.;
RT "Structure of a linear plasmid of the yeast Kluyveromyces lactis;
RT compact organization of the killer genome.";
RT Curr. Genet. 9:147-153(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84297209; PubMed=6473099;
RA Stark M.J.R., Mileham A.J., Romanos M.A., Boyd A.;
RT "Nucleotide sequence and transcription analysis of a linear DNA
RT plasmid associated with the killer character of the yeast
RT Kluyveromyces lactis.";
RL Nucleic Acids Res. 12:6011-6030(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85037931; PubMed=6387625;
RA Tokunaga M., Kawamura A., Hishinuma F.;
RT "Cloning and nucleotide sequences of the linear DNA killer plasmids
RT from yeast.";
RL Nucleic Acids Res. 12:7581-7597(1984).
RN [4]
RP IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.
RX MEDLINE=87004569; PubMed=3758030;
RA Stark M.J.R., Boyd A.;
RT "The killer toxin of Kluyveromyces lactis: characterization of the
RT toxin subunits and identification of the genes which encode them.";
RL EMBO J. 5:1995-2002(1986).
RN [5]
RP SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.
RX MEDLINE=90259069; PubMed=2342564;
RA Bradshaw H.D. Jr.;
RT "Killer toxins.";
RL Nature 345:299-299(1990).
RN [6]
RP CHITINASE ACTIVITY OF ALPHA-SUBUNIT.
RX MEDLINE=91301161; PubMed=2070799;
RA Butler A.R., O'Donnell R.W., Martin V.J., Gooday G.W., Stark M.J.R.;
RT "Kluyveromyces lactis toxin has an essential chitinase activity.";
RL Eur. J. Biochem. 199:483-488(1991).
CC -1- FUNCTION: THE ALPHA SUBUNIT IS A POTENT EXOCHITINASE. ALONG WITH
CC THE BETA SUBUNIT IT PLAYS A ROLE IN THE INITIAL INTERACTION OF THE
CC TOXIN WITH SENSITIVE CELLS AND ALLOW THE GAMMA SUBUNIT (THE ACTIVE
CC TOXIN) TO GAIN ENTRY INTO THE CELL.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SUBUNIT: THE KILLER TOXIN IS COMPOSED OF THREE SUBUNITS: ALPHA,
CC BETA AND GAMMA.
CC -1- PTM: RF2 IS POTENTIALLY SPLIT BY MEMBRANE-BOUND BASIC AMINO ACID-
CC SPECIFIC PEPTIDASE TO YIELD THE ALPHA AND BETA SUBUNITS.
CC -1- SIMILARITY: LOCAL, TO OTHER CHITIN-BINDING PROTEINS.
CC -1- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC ----- X07127; CAA30137.1; -
DR EMBL; X00762; CAA25334.1; -
DR EMBL; X01095; CAA25569.1; -
DR PIR; S07915; S07915.
DR HSSP; P02877; 1HEV.
DR INTERPRO; IPR001002; -
DR INTERPRO; IPR001223; -
DR INTERPRO; IPR001579; -
DR PFAM; PF00704; Glyco_hydro_18; 1.
DR PRINTS; PR00187; chitin_binding; 1.
DR PROSITE; PS00026; CHITIN_BINDING.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Plasmid; Toxin; Glycoprotein; Signal; Hydrolase; Glycosidase;
KW Chitin degradation.
FT SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 29 POTENTIAL.
FT CHAIN 30 892 KILLER TOXIN ALPHA SUBUNIT (PROBABLE).
FT CHAIN 895 1146 KILLER TOXIN BETA SUBUNIT (PROBABLE).
FT DOMAIN 310 363 CHITIN-BINDING (BY SIMILARITY).
FT DOMAIN 364 635 CATALYTIC.
FT ACT_SITE 495 495 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 771 771 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 858 858 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 868 868 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 876 876 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 995 995 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1082 1082 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1117 1117 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 1146 AA; 128936 MW; BF4B1764EB465DE6 CRC64;

Query Match 6.8%; Score 90.5; DB 1; Length 1146;
Best Local Similarity 19.3%; Pred. No. 7.8;
Matches 69; Conservative 34; Mismatches 104; Indels 151; Gaps 19;

QY 23 ESLENLKDKNVENIWL-ILENRSF-DNII- - - - -GGVRRQGLDNPINNGPFCNRYKNSADP 75
DB 76 ESFYNSNGNYGPMVGSVNRGFKDNLIPMEVKNVGIPTL- - - - -YLEYDGGGDP 131
QY 76 - - - - -SSGKYCTQAKDYDSVNDPDHSVTVG- - - - -NNLEFYG 107
DB 132 MKSRGIIIDTTSRDVTVKAAKLWSQKKLSYEGSKNYQATCYLSYAYRKPIVNDNFVG 191
QY 108 T- - - - -YTPNNGAIGASGVVADQSGFLNAQLNDY- - - - -PKLAP- 141
DB 192 TCDYFT- - - - -LESKTPADQSGINGESLQYGNPNLDFSKLSAQPICTIGNPNTKPS 246
QY 142 - - - - -EATROVMGYITEEVETLVLDVDEFTFN-SWFSC- - - - - 176
DB 247 KNSGCSCKTYKVSSESCSSTAVKIY- - - - -PLSUNDIENYNKGNYGWKGCSLQKQDYNL 301
QY 177 - - - - -VPGP-TNP- - - - -NRLCALAG- - - - -TAAGHG 197
DB 302 CVSDGSAPRPVSNPIAECGLPAGCKYNKACPLNACCSEFGCLTKDYCKKSSITGAP 361
QY 198 KNDODFLNYCISSKSIFEAANEKGV-SWLYDQ- - - - -TNGEFEPDSLFFTYVN 245
DB 362 GTDGCFSNGCYGSTNVKSTFKKIAYWLDKADKLPKIPNGPY- - - - -DILHAFVN 417

RESULT 6
ID D29B_ARATH STANDARD; PRT; 584 AA.
AC Q06737.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE DESICCATION-RESPONSIVE PROTEIN 29B.
GN RD29B.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

```

```

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA MEDLINE-94143472; PubMed-8310052;
RX Yamaguchi-Shinozaki K., Shinozaki K.;
RT "Arabidopsis DNA encoding two desiccation-responsive rd29 genes.";
RL Plant Physiol. 101:1119-1120(1993).
RN [2]
RP SEQUENCE FROM N.A., AND INDUCTION.
RC STRAIN-CV. COLUMBIA;
RA MEDLINE-93173109; PubMed-8437577;
RX Yamaguchi-Shinozaki K., Shinozaki K.;
RT "Characterization of the expression of a desiccation-responsive rd29
gene of Arabidopsis thaliana and analysis of its promoter in
transgenic plants.";
RL Mol. Gen. Genet. 236:331-340(1993).
CC -!- INDUCTION: BY ABSICISIC ACID (ABA) AND WATER STRESS.
CC -!- SIMILARITY: BELONGS TO THE LT178/RD29A; RD29B; LT165 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13044; BAA02375.1; -.
KW Multigene family.
SQ SEQUENCE 584 AA; 62710 MW; 7972B6526A2D01F6 CRC64;

Query Match 6.8%; Score 90; DB 1; Length 584;
Best Local Similarity 22.7%; Pred. No. 3.6;
Matches 59; Conservative 29; Mismatches 92; Indels 76; Gaps 13;

QY 4 TSEYTSVRAPGYPGKPGSK- - - - -ESLENLKDKNVENIWLILE- - - - -NRSP- - - - -DNI 48
DB 128 TSDWEAKRAPHYPLGVSEFSDFSGESREAHQEPILNTPVSLLSATSDVTRTFAPGGEDDY 187
QY 49 LGGVRRQGLDNP- - - - -INNGPFCNRYKNSADPSSGKYCTQAKDYDSVNDPDHSVTVGNLEFY 106
DB 188 LGGORKVNVETPKRLEEDP- - - - -AAPGGSDYLSGVSNYQSKVTDPTKKEAG- - - - - 235
QY 107 GTYTPNNGAIGASGVVADQSGFLNAQLNDYPKLAPERATROVMGYITEEVETLV- - - - -VDL 163
DB 236 - - - - -VPETAEBSLGRMKVTDSE- - - - -PDQKSRQ- - - - -GREEDFPTRSHFDL 274
QY 164 VDEFTTFNSWFCVPGPTNPRLCALAG- - - - -TAAGHGKNDODFLNYGISSKSIFEAANEKGV 222
DB 275 KKE- - - - -SDLNKNSPARFGESKAG- - - - -MEEDFPTRG- - - - -DYKVESGL 312
QY 223 SWLYDGTNGEFPD 237
DB 313 GRDLPTGTHDQFSPE 327

RESULT 7
ID LT65_ARATH STANDARD; PRT; 600 AA.
AC Q04980;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE LOW-TEMPERATURE-INDUCED 65 KDA PROTEIN.
GN LT165.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
RN [1]

```



RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA; TISSUE=LEAF;  
 RX MEDLINE=93192524; PubMed=8448363;  
 RA Nordin K., Vahala T., Palva E.T.;  
 RT "Differential expression of two related, low-temperature-induced  
 genes in Arabidopsis thaliana (L.) Heynh. ";  
 RL Plant Mol. Biol. 21:641-653(1993).  
 CC -!- INDUCTION: BY LOW TEMPERATURE, AND MOSTLY BY WATER STRESS OR  
 CC ASCISIC ACID (ABA).  
 CC -!- SIMILARITY: BELONGS TO THE LTI78/RD29A; RD29B; LTI65 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X67670; CAA47902.1; -;  
 DR Multigene family; Repeat.  
 KW DOMAIN 24 27 POLY-GLU.  
 FT DOMAIN 64 67 POLY-ASP.  
 FT DOMAIN 402 509 5 X 5 AA REPEATS OF [IV]-[AMS]-[EST]-  
 FT K-L.  
 FT REPEAT 402 406 1.  
 FT REPEAT 440 444 2.  
 FT REPEAT 458 462 3.  
 FT REPEAT 488 492 4.  
 FT REPEAT 505 509 5.  
 FT DOMAIN 512 515 POLY-GLY.  
 FT SEQUENCE 600 AA; 64510 MW; 9569296786D52570 CRC64;

Query Match 6.8%; Score 90; DB 1; Length 600;  
 Best Local Similarity 22.7%; Pred. No. 3.7;  
 Matches 58; Conservative 29; Mismatches 92; Indels 76; Gaps 13;  
 QY 4 TSEYTSVREAPGYKPGSK-----ESIEIENLKDVENIVLILE---NRSP-----DNI 48  
 Db 155 TSDWEAKREAPHYPLGVSEFSDRGESRAHQEPLNTPVLSLSTEDVTRTAPGEGDY 214  
 QY 49 LGGVRRGLDNP--INNGPFCNKNASDPSSGKYCTQAKYDVSFNDPDSHVSYTNLEFY 106  
 Db 215 LGGQRKVNVEPKRLEEDP-----AAPGGSDYLSGVSNYSQSKVDPTTHKEAG----- 262  
 QY 107 GTYTPNNGAIGSKVVDQSGFLNQLNDYPKLAPEEATRVGMGYTVEEVPTL---VDL 163  
 Db 263 ---VPEAESLGRMKVTDSE-----PDQKSRQ---GREEDFTSRSHFDL 301  
 QY 164 VDEFTTNFWSFCVPGPNPRLCALAG-TAAGHGKNDDDLFLNYGSSKSIPEANEGV 222  
 Db 302 KKE-----SDINKNSPARGESKAG---MEEDPPTRG-----DVKVESGL 339  
 QY 223 SWLNYDGTNGEFPD 237  
 Db 340 GRDLPTGTHDQESPE 354

RESULT 8  
 GLN3\_LUPAN STANDARD; PRT; 315 AA.  
 AC P14636;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUTAMINE SYNTHETASE NODULE ISOZYME (EC 6.3.1.2) (GLUTAMATE--AMMONIA  
 DE LIGASE) (FRAGMENT).  
 OS Lupinus angustifolius (Narrow-leaved blue lupine).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Lupinus.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. UNTHARVEST; TISSUE=ROOT NODULES;  
 RX MEDLINE=91370830; PubMed=2577495;  
 RA Grant M.R., Carne A., Hill D.F., Farnden K.J.F.;  
 RT "The isolation and characterization of a cDNA clone encoding Lupinus  
 angustifolius root nodule glutamine synthetase. ";  
 RL Plant Mol. Biol. 13:481-490(1989).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) -> ADP + GLUTAMINE +  
 CC ORTHOPHOSPHATE.  
 CC -!- SUBUNIT: HOMOOCTAMER.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X15578; CAA33605.1; -;  
 DR PIR: S07462; AJYLNH  
 DR INTERPRO: IPR001691; -;  
 DR PFAM: PF00120; gln-synt; 1.  
 DR PROSITE: PS00180; GLNA\_1; 1.  
 DR PROSITE: PS00181; GLNA\_ATP; 1.  
 KW Nitrogen fixation; Ligase; Multigene family.  
 FT NON\_TER 315 315  
 SQ SEQUENCE 315 AA; 34441 MW; DC3B61A903F442F3 CRC64;

Query Match 6.7%; Score 88.5; DB 1; Length 315;  
 Best Local Similarity 25.0%; Pred. No. 2.2;  
 Matches 58; Conservative 21; Mismatches 84; Indels 69; Gaps 15;  
 QY 27 NLKDKVENI---VMLILENSFONILGG-----VRRQGLDNPING---PFCNKNAS 73  
 Db 11 NLSDTTEKIAYEW-----VGGSGVDLSKARTLSGVPNDPSKLPKNWY---- 55  
 QY 74 DPSSGKYCTQAKDYDS-----VFNDPDHVSYTN--LEFYGTYPNNGATASGVVAD 124  
 Db 56 ---DGSSTGQAPKQSEVILWPQAFIKDPFR--RGNILVICDTYTP-----SGKPIPT 104  
 QY 125 QSGFLNQLNDYPKLAPEEATRVGMGYTVEEVPTLVDLVEFTTNSFSCVPGPTNPN 184  
 Db 105 NKRAAAKIFSHPDVAEEP-----WFGIEQEYTLQ-KDIHWPICMALGGPGPGP- 156  
 QY 185 RLCALACTAAGHGKNDDDLFLNYGSSKSIPEANEGVSWLNYDGTNGEFP 236  
 Db 157 YYCG-TGAEKAFGRDIVD-----SHYKACLYAGI---NISGINAEVMP 195

RESULT 9  
 DLG1\_DROME STANDARD; PRT; 960 AA.  
 ID DLG1\_DROME  
 AC P31007;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE LETHAL(1)DISCS LARGE-1 TUMOR SUPPRESSOR PROTEIN.  
 GN DLG1.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91330294; PubMed=1651169;  
 RA Woods D.F., Bryant P.J.;  
 RT "The discs-large tumor suppressor gene of Drosophila encodes a  
 RT guanylate kinase homolog localized at septate junctions. ";  
 RL Cell 66:451-464(1991).





QY 72 ASDPSGKYCTQAKDYD---SVFND-----PDHSTVGN-----NLEFGVGYTP-- 111  
 Db 101 AIDPD-----KDHECMKIFSDAGIYIVADLSEPTVSINRNPNWNLDKYRTKVI 152  
 QY 112 -----NN--GATASGVV-----ADSGFLNAQLNDYFKLAPEEATROV-MGYVTEEEV 157  
 Db 153 DKMQEYSNVLGFFAGNEVYNNRSNTDASAFVKAIRDKMKYIKESDYRQIPGVSYSSNDE 212  
 QY 158 PLYDLVDEFTNSWFSVCPGPTNPNRLCALAGTAAGKND--DFLNYGTS-----S 210  
 Db 213 EIRVAIAD-----YFSC-----GSLDDRADF--FGINMYEWCG 243  
 QY 211 KSIFEAANEK 220  
 Db 244 KSTFETSGYK 253

RESULT 13  
 YFBL\_ECOLI STANDARD; PRT; 323 AA.  
 AC F76482;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL 35.9 KDA PROTEIN IN ELAD-NUON INTERGENIC REGION.  
 GN YFBL.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000317; AAC75331.1; ALT\_INIT.  
 DR ECOGENE: EG14096; yfbl  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 4 24 POTENTIAL.  
 SQ SEQUENCE 323 AA; 35931 MW; B9DB5AD1289C6013 CRC64;

Query Match 6.4%; Score 84.5; DB 1; Length 323;  
 Best Local Similarity 26.6%; Pred. No. 4.9;  
 Matches 42; Conservative 17; Mismatches 68; Indels 31; Gaps 6;  
 QY 1 SPVTSVTSVREAPPYKPGSKESIEMLKDKVENIVMLLENRSFDNLGGVRRGRLDNP 60  
 Db 37 SPEQLEKT-VRYLTQTVHPRSDNIDNLSAEYIKYEVFSS-----GARVTSQDVP 87  
 QY 61 INNGPFCNYKNSDSSSGKYCTQAKDYDSVFNPDHSTVGNLEFYGTVPNNGAASGK 120  
 Db 88 ITGGPKNYVADYGPADGPLIIICAHYDSA-----SSYENDQL-----TTPGADDNASGV 138  
 QY 121 VVAQSGFLNAQLNDYFKLAPEEATR-----QVMGYTTEE 155  
 Db 139 -----AGLELRLHQQVPKGTGVOLVAYASEE 166

RESULT 14

GLN2\_VITVI  
 ID GLN2\_VITVI STANDARD; PRT; 356 AA.  
 AC P51119;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE GLUTAMINE SYNTHETASE CYTOSOLIC ISOZYME 2 (EC 6.3.1.2) (GLUTAMATE--  
 DE AMMONIA LIGASE).  
 GN GSI-2 OR GSI;2.  
 OS Vitis vinifera (Grape).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. SULTANINA; TISSUE=SHOOT;  
 RX MEDLINE=97000911; PubMed=8843941;  
 RA Loulakis K.A., Roubelakis-Angelakis K.A.;  
 RT "Characterization of Vitis vinifera L. glutamine synthetase and  
 RL molecular cloning of cDNAs for the cytosolic enzyme.";  
 CC Plant Mol. Biol. 31:983-992(1996).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) -> ADP + GLUTAMINE +  
 CC ORTHOPHOSPHATE.  
 CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X94321; CAA63982.1; -.  
 DR INTERPRO: IPR001691; -.  
 DR PFAM: PF00120; gln-synt; 1.  
 DR PROSITE: PS00180; GLNA\_1; 1.  
 DR PROSITE: PS00181; GLNA\_ATP; 1.  
 KW Ligase; Multigene family.  
 SQ SEQUENCE 356 AA; 39322 MW; 5EB0D289F971E5D7 CRC64;  
 Query Match 6.4%; Score 84.5; DB 1; Length 356;  
 Best Local Similarity 24.1%; Pred. No. 5.5;  
 Matches 57; Conservative 24; Mismatches 77; Indels 79; Gaps 16;  
 QY 27 NLKDKVENI-----VMLLENRSFNTLGG-----VRRQGLDNPING---PFCNKNAS 73  
 Db 11 NLSVTEKIIAEYIW-----IGSGMDLRSKARTLSGVSVDPHKLPKNY----- 55  
 QY 74 DPSSGKYCTQAKDYDS-----VFNDPDHSTVGN--LEFYGTVPNNGAASGVVAD 124  
 Db 56 ---DGSSTGQAPCEDSEVILYPOAIEKDPFR--RCNNILVMCDITYP-----ACEIPT 104  
 QY 125 QSGFLNAQLNDYFKLAPEEATROVMGYTTEEVPLVDLVDVDEFTTNSW-----FSCVPG 179  
 Db 105 NKRNRKAKFSFPEVLAET-----WYGEQETTLQNSV-----KWPICGWVGYGYPG 152  
 QY 180 PTPNRLCALAGTAAGHKNDLDFLNYGISKSIKSEIFEAANEKGVSWLNYDGTGTEPEP 236  
 Db 153 PQGP-YVCGI-GADKAFGRDIVD-----SHYKACLYAGI---NISSINGEVP 195

RESULT 15  
 GPC4\_MOUSE  
 ID GPC4\_MOUSE STANDARD; PRT; 557 AA.  
 AC P51655;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE GLYPICAN-4 (K-GLYPICAN).  
 GN GPC4.

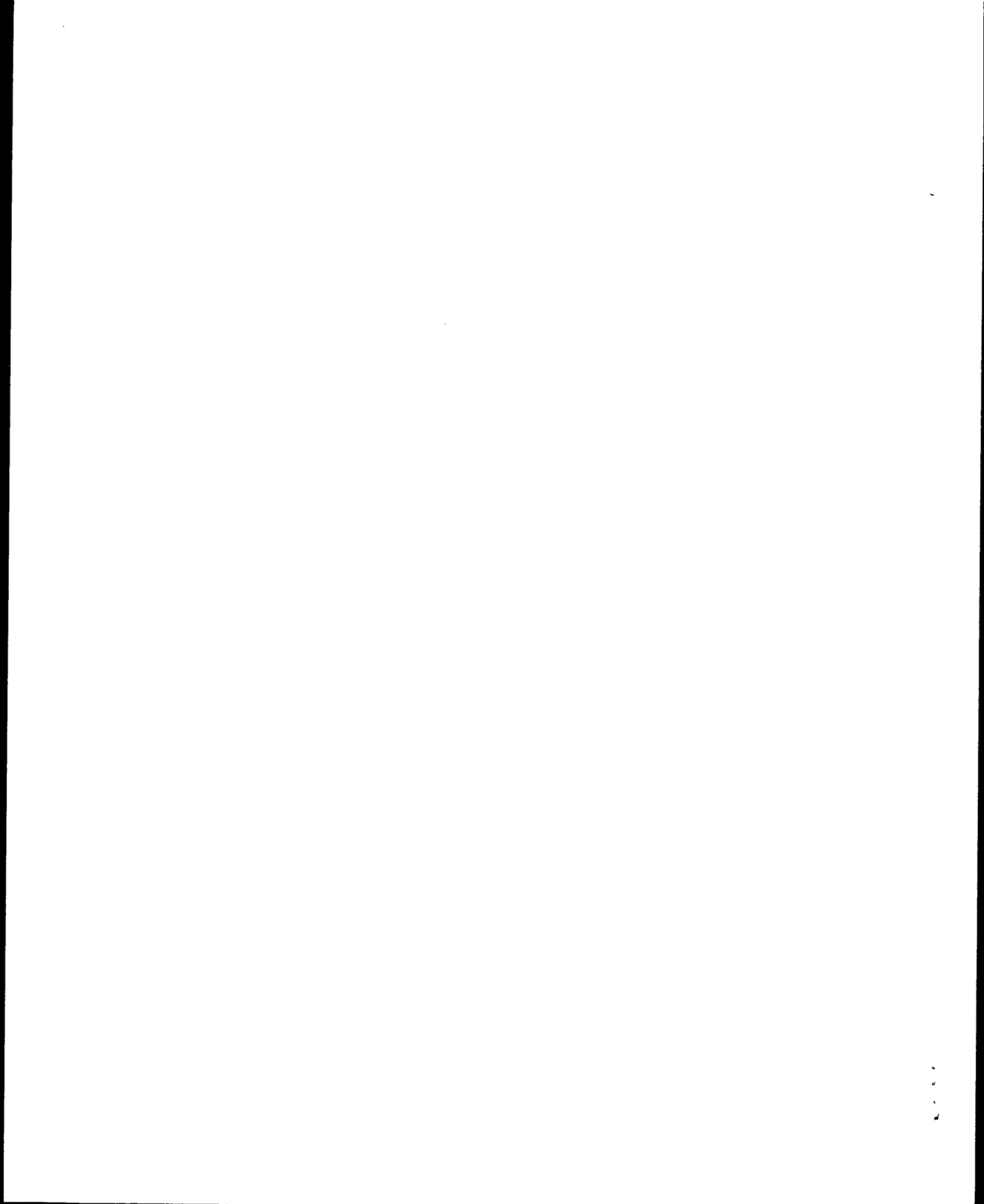
Tue Feb 27 08:39:08 2001

us-09-426-072-2\_copy\_20\_264.rsp

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BA1B/C;  
RX MEDLINE=95386580; PubMed=7657705;  
RA Watanabe K., Yamada H., Yamaguchi Y.;  
RT "k-glypican: a novel GPI-anchored heparan sulfate proteoglycan that  
is highly expressed in developing brain and kidney.";  
RL J. Cell Biol. 130:1207-1218(1995).  
CC -!- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN. MAY BE  
INVOLVED IN THE DEVELOPMENT OF KIDNEY TUBULES AND OF THE CENTRAL  
NERVOUS SYSTEM.  
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN DEVELOPING BRAIN AND  
KIDNEY.  
CC -!- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; X83577; CAA58560.1; -.  
DR MGD; MGI:104902; GPC4.  
DR INTERPRO: IPR001863; -.  
DR PFAM: PF01153; Glypican; 1.  
DR PROSITE: PS01207; GLYPICAN; 1.  
KW Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.  
FT SIGNAL 1 18  
FT CHAIN 19 529  
FT PROPEP 530 557  
FT LIPID 529 529  
FT CARBOHYD 514 514  
FT CARBOHYD 494 494  
FT CARBOHYD 498 498  
FT CARBOHYD 500 500  
FT SEQUENCE 557 AA; 62616 MW; F216AA0FF45F6840 CRC64;  
OY 47 NILGG-VRRQG-LDNPIIN-----GPGCNKNASDPSSGKYCTQAKDYDSVF 91  
DB 272 NMRGCLANQGLDFEWNFFIDAMLMVAERLEGPF-NIESVMDP-----IDVKISDAIM 324  
OY 92 NDPDHSVTGNLLEFYGTTPNGAIASGV--VADQSGFLNALNDYKLAPEATRQVM 149  
DB 325 NMQDMSVQVSQKVFQCGCPK--PLPAGRISRSISAF-SARFRPY---HPEQRPTAA 378  
OY 150 GYVTEEEPTLVLDVDEFTTFNSWFCVPGTPNPNRLCALAGTAAGHGKNDDDLNYGIS 209  
DB 379 GTSLDRLVTDVKEKLKQKKE--WSSL-----PSTVNDERMAAGN-ENEDDCWNGKG 429  
OY 210 SKSIFEAANEKGVSWLNYDGTNGEFEPDS 238  
DB 430 SRVLFVGTG-----NGLANQGNPNPEVQVDT 454

Query Match 6.4%; Score 84.5; DB 1; Length 557;  
Best Local Similarity 24.9%; Pred. No. 9.9;  
Matches 52; Conservative 31; Mismatches 83; Indels 43; Gaps 13;

Search completed: February 26, 2001, 19:00:18  
Job time: 17412 sec









```

15 PheGlyTyrLysProGlySerLysGluSerIleGlu..... 26
16687 TTTGCTACTTACTAATTCATCTCATCTCATTCATGATTCATCAAACTCCCA 16736
27 .....AsnLeuLysAspLysValGluAsnIleValTrpLeuIleL 40
16737 GAAACCCCAAGATCCCAAGGACCAATCAAGACTATAGTCGTTGTATGCA 16786
40 euGluAsnArgSerPheAspAsnIleLeuGlyGlyValArg.....Arg 54
16787 TGGAGATCGCTCTTCATCATCTCGGTGCTCAATCGACCCGA 16836
55 GlnGlyLeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrLysAs 71
16837 CCCGAATCGACGGA.....TTAACGGCAAGAGTCAATCCCTCCAA 16880
71 nAlaserAspProSerSerGlyLysTyrCysThrGlnAlaLysAspTyrA 88
16881 CGTCTCGATCCAAATTCACAAAGATCTTCGTCACAG.....G 16921
88 spSerValPhe.....AsnAspProAspHisSerValThrGlyAsnAsn 102
16922 ACGCCGCTTCCTAGACATCGATCCGGTCACTCATTCCAAGCGATCCGA 16971
103 LeuGluPheTyrGlyThrTyrThrProAsnAsnGlyAlaIleAlaSerG 119
16972 GAACAGATATCGGG.....TCTACGATACATCCGGCGACCC 17009
119 yLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLeuAsnAsp 136
17010 GAAATG.....AACGATTCCGACAGCAATCAGAGATATGG 17047
136 yrProLysLeuAlaProGluAlaThrArgGlnValMetGlyTyrTyr 152
17048 AGCTGGAATGGCT.....AAGAACGTTATGAGCGGGTTC 17082
153 ThrGluGluValProThrLeuValAspLeuValAspGluPheThrH 169
17083 AAACCGGATCTTACCGGTTACCGGATAGCGAACGAGTTCGGAGT 17132
169 rPheAsnSerTrpPheSerCysValProGlyProThrAsnProAsnArg 186
17133 TTTGATCGATGGTTCGCTCAGTACCAACTCGACTCAACCGAATCGGT 17182
186 euCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyLysAsnAspAsp 202
17183 TCTACGTTCACTAGCTACTTCTCATGGATGCTCAAGCAACGTGAGAAA 17232
203 PheLeuAsnTyrGlyIleSerSerLysSerIlePheGluAlaAlaAsn 219
17233 GATCTCGTTAAAGGGTTCTCTCAAAAGACGATCTTTGATTCGTTAGATGA 17282
219 uLysGlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGlu 236
17283 GAACGGACTCAGTTT.....GGGATTTATTACCAAAACATCC 17320
236 toAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSerAsnVal 252
17321 CTGGGAGCTTCTTCTCAAGTCTCTCGTAGGCTT.....AAGCATTTG 17364
253 ValProValGluAsn.....PhePheGlnAspAlaTyrLeuG 265
17365 GTGAATTCCTAGCTATGCTCACTTAAGTTCAGCTCGACGCTAGCTTGG 17414
265 yValLeuProLysPheSerTyrIleAsnProSerCysCysGlyThrAsn 281
17415 GAAGCTTCCCAATTTATCTGCTGGAACAGAGATATTCGATATCGCAT 17464
282 .....ThrAsnSerMetHisProThrGlyAsnValSerTyrGlyGlu 295
17465 TTTTTCGGCTATGATGATCATCGTCGATGACGTGCGCGGGGCA 17514
296 ValPheValLysGlnIleTyrAspAlaIleArgGlnGlyProGlnTrpAs 312

```

```

17515 CGGTTTGAAGGAAGTTTATGAGAGCTTACGAGTAGTCCACAGTGGAA 17564
312 pLysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheTyrAspH 329
17565 AGAGATGGCTTTCCTTACTTATGATGAGCAGCGTGGTGGTTTATGACC 17614
329 isValProProLeu.....AlaValArgProAspAsnLeuThrTyr 343
17615 ATGTTCCACGCGGTTAAAGGTGTCCTAACCTGATGGTATCAIT... 17661
344 ThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheAspArgLe 360
17662 .....GGACCTGATCCGTTTACTTTGGTTTGACCGATT 17696
360 uGlyGlyArgMetProThrTrpValIleSerProTyrSerLysLysGly 376
17697 GGGTGTTCGGTCCCTACCTTCTGATCTCTCTCTGGATTGAGAAGGCG 17745

seq_name: gb_p12:ATT8P19

seq_documentation_block:
LOCUS   ATT8P19           96679 bp    DNA             PLN       02-DEC-1999
DEFINITION   Arabidopsis thaliana DNA chromosome 3, BAC clone T8P19.
ACCESSION   AL133315
VERSION     AL133315.1   GI:5523080
KEYWORDS
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE   1 (bases 1 to 96679)
AUTHORS     Cholsne,N., Robert,C., Brottier,P., Wincker,P., Cattolico,L.,
            Artiguenave,F., Saurin,W., Weissenbach,J., Mewes,H.W., Lemcke,K.,
            Mayer,K.F.X., Quetier,F. and Salanoubat,M.
            Unpublished
JOURNAL     EU Arabidopsis sequencing,project.
REFERENCE   2 (bases 1 to 96679)
AUTHORS     Submitted (02-DEC-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,Project
            Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
            d'Interet Public, Centre National de Sequençage - GENOSCOPE; 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France;
            http://www.genoscope.cns.fr
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.

FEATURES             Location/Qualifiers
     source            1..96679
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
     gene              36..963
                     /gene="T8P19.10"
     exon              36..86
                     /gene="T8P19.10"
                     /number=1
     CDS                join(<36..86,324..475,546..963)
                     /gene="T8P19.10"
                     /codon_start=1
                     /product="hypothetical protein (fragment)"
                     /protein_id="CA62339.1"
                     /db_xref="GI:5523081"
                     /translation="KTSVKELDLEALTEROHKHLMMKARGCYKIDKLRRNIEM
                     DEYDFLWRSLEEREALLRISRQALGLEPREGYKPGSGFGDKQDPTPSALYQY
                     DWGEFANSEISKQRMKAHNSIVKGNWVMDYDDAIKQTIETREKRGSTLASQE
                     ETSESENEEDDDDFDDFDYDSILSDESSIGYSEQQPLVNGTQVLTD"
     intron             87..323

```





Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr.  
Rockville, MD 20850, USA  
e-mail: xlin@tigr.org  
BAC clone T21P5 is from Arabidopsis chromosome III and is near the  
molecular marker ml172.  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.

Genes were identified by a combination of three methods: Gene  
prediction programs including GRAIL (available by anonymous ftp  
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of  
Washington), Genscan (Chris Burge, and NetplantGene  
(http://www.cbs.dtu.dk/netgene/cbsnetgene.html), and searches of the  
complete sequence against a peptide database and the Arabidopsis  
EST database at TIGR (http://www.tigr.org/tdb/at/at.html).  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as 'unknown' proteins.  
Genes without protein or EST similarity, that are predicted by more  
than two gene prediction programs over most of their length are  
annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
annotated as 'tRNAscan-SE' (Sean Eddy,  
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are  
identified by repeatmasker (Arian Smit,  
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of  
genomic sequence that are not annotated as genes but have predicted  
exons by GRAIL are annotated as misc features.

#### FEATURES

```

source
    location/Qualifiers
1..82289
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="III"
/map="ml172"
/clone="T21P5"

misc_feature
1..24217
/feature="Overlap with BAC clone T12J13
(AC009327:65718..89934)."
complement(join(<8..215,451..587,675..857,959..1050,
1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
2979..3100))
/feature="T21P5.1"
complement(<8..>3100)
/feature="T21P5.1"
/feature="T21P5.1"
/feature="similar to hypothetical protein GB:CA838918
[Arabidopsis thaliana]"
complement(join(8..215,451..587,675..857,959..1050,
1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
2979..3100))
/feature="T21P5.1"
/feature="unknown protein"
/codon_start=1
/protein_id="AAF01600.1"
/db_xref="GI:6017117"

translation="MSQQLSGNNIPLESVYSLVKNADKKFKIRDLFFPERSRYEN
IFKVFVYQLWKQFQENKQKVEAGLRKWEIGIASRTAQLYIGHYMRTSDAGYLS
ESYVFFVLTREYEDKGLFDQDLNANKQLRFARFLMCLVLGRREYHQLVDQFRK
LIDCKRTFOETDFKEMVVAQETVRFKSDTAFNNIPLRLYSLVLDPLNDACTPRAS
RSLRDLTALISYVCNEVYSELTLDLFRMLQCLEWEPGSLYQSTGAKMGONAPGVG
ARINSQMDLPTPPNPKAVLYRPSITHTFLAVLATICEELPSHGILLLYLSASGKIG
QISSPLSARSATSVENILRDPSHTIKQETEPSLQITPGSSYIYPSDLVPTKRP
LFIIIDSDSTVFKNICCAEKGEPALLISPSHTPTPLISADFSRPGSLFIILTSP
VOACLLISASDMETDIFKAKLLSSMNEWASILASDTLHFVWSQILLKDFPLR
RLLRFFTCRAVLALITVFVNNKQFCCPSLPSLPTAPAVASVQMANVFGAT
SKFTIPODITPLESF"
join(<3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)
/feature="T21P5.2"
<3627..>6144

mRNA
join(<3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)

gene
/feature="T21P5.2"
<3627..>6144

```

```

/feature="T21P5.2"
/feature="predicted by genscan"
join(3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)
/feature="T21P5.2"
/feature="hypothetical protein"
/codon_start=1
/protein_id="AAF01580.1"
/db_xref="GI:6017097"

translation="MDDNRSSSEIKRHETEKTIASRKLEDNTKLIQDPEEMALYAK
VRSDEEHSIQERIAAACLDQMQLNLEKGLERKCADLRVAIDKONESVTSALNEL
ARRGDLEENLAHLDKJEDERYIFMTSLGLLAEGYVWPVRVANATISGICKHLH
DQLQWTKACNDIRIELSSIVENQPGDFISLKDNDHPRNSKTAQSYSGYDNGDYOTN
BQLLPMMENVTNPMYINMOTLESRNNOIGESQSGQIFPOPKRENFYPLSGVAGKE
MIQREKASNSMDAYNGNEEFASHVEEGPGFIIDGDAICERKVLGGCPVR
ETTLICFQWRILEDGTQRYLEGATHEPXLVTADDVDKLIAVEICPMDDQGRQVKYRD
ESGIYSFNSYVSKDVLIMQELVRLFANDQNKIRCGNFWGI"
6225..6289
/feature="exon predicted by xgrail, quality marginal"
6384..6503
/feature="exon predicted by xgrail, quality excellent"
6794..6839
/feature="exon predicted by xgrail, quality excellent"
6928..7033
/feature="exon predicted by xgrail, quality marginal"
7293..7411
/rpt_family="(TAAAA)n"
complement(7488..7547)
/feature="exon predicted by xgrail, quality
excellent_shadowexon"
complement(7654..7784)
/rpt_family="(TAA)n"
complement(7819..7864)
/feature="exon predicted by xgrail, quality good_shadowexon"
8054..8113
/feature="exon predicted by xgrail, quality marginal"
8263..8380
/rpt_family="(TAAAA)n"
complement(8592..8627)
/feature="exon predicted by xgrail, quality
marginal_shadowexon"
<8652..>9527
/feature="T21P5.3"
<8652..>9527
/feature="T21P5.3"
/feature="contains zinc finger motif, C3HC4 type (RING
finger)"
8652..9527
/feature="T21P5.3"
/feature="unknown protein"
/codon_start=1
/protein_id="AAF01602.1"
/db_xref="GI:6017119"

translation="MGSTGNPNWGTYYSDYSCSQGVYCPQWCYVIFPPPPSFY
LDDDDSSSDSPLLIAGILASAFILVSYTILSKYCHRRRRSSSTSAAINRI
SSDYTWGNNNNNGATNPNOTIGGGGGLDESLIKSTIVTKYKMGDFVSSDGS
VCLSEFQENSLRLPLKCNHAFVPCIDTWLKSHSNCPLCRAFITSSAVEIVDLTQ
QIVTENNISTGDDSVVYVNLDLNLSRNETVNEGSTPTXGDAGFKRRRAIGVYK
LRRRCGIDRRRYTKGD"
complement(10838..10896)
/feature="exon predicted by xgrail, quality
marginal_shadowexon"
11174..11254
/feature="exon predicted by xgrail, quality good"
11690..11700
/feature="exon predicted by xgrail, quality marginal"
join(<11814..12256,12381..12578,12676..13085,13162..13332,
13418..>13761)
/feature="T21P5.4"
<11814..>13761
/feature="T21P5.4"
/feature="similar to unknown protein GB:AAC32238 [Arabidopsis

```



13083 CAGGTAAGTATGCTGAGACAATAATTTAATTATACATAATGCTGTTAAT 13132  
 384 .....AspProValThrGlyLysProAlap 392  
 13133 TAAATGACAAATTTTGTAAATGCTGATGCTTCATGCTGCAATGGC 13182  
 392 ro.....TyrSerAlaThrSerValLeuLysThrLeuGly 403  
 13183 CGTACCAATATGTCACAAATGAGCATTCGTCATCCCGACGGTCAAG 13232  
 404 TyrLeuTrpAspIleGluAspPhe...ThrProArgValAlaHisSerPr 419  
 13233 AAAATTTCAACATCAAGATTTTTCACAAGAGAGATTCGTGGCTGG 13282  
 419 oSerPheAspHisLeulle...GlyThrThrLeuArgGluAspAlaPro 434  
 13283 CACTTTGAAATCTGTTATTACGAGAACTACCAAGACAAGATTCGCT 13331  
 seq\_name: gb\_pl2:ATAC009327  
 seq\_documentation\_block:  
 LOCUS ATAC009327 89934 bp DNA PLN 21-OCT-1999  
 DEFINITION Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence,  
 complete sequence.  
 ACCESSION AC009327  
 VERSION AC009327.6 GI:6091751  
 KEYWORDS HTG.  
 SOURCE  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 Brassicales; Brassicaceae; Arabidopsids.  
 REFERENCE 1 (bases 1 to 89934)  
 Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.,  
 Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E.,  
 Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.  
 Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence  
 2 (bases 1 to 89934)  
 Lin, X. and Kaul, S.  
 Direct Submission  
 Unpublished  
 JOURNAL  
 REFERENCE 3  
 Lin, X.  
 Submitted (16-AUG-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org  
 3 (bases 1 to 89934)  
 Lin, X.  
 Direct Submission  
 Submitted (21-OCT-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 On Oct 21, 1999 this sequence version replaced gi:5868934.  
 Address all correspondence to:  
 Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr.  
 Rockville, MD 20850, USA  
 e-mail: xlin@tigr.org  
 BAC clone T12J13 is from Arabidopsis chromosome III and is near the  
 molecular marker ml72.  
 The orientation of the sequence is from SP6 to T7 end of the BAC  
 clone.  
 Genes were identified by a combination of three methods: Gene  
 prediction programs including GRAIL (available by anonymous ftp  
 from [arthur.epm.ornl.gov](http://arthur.epm.ornl.gov)), Genefinder (Phil Green, University of  
 Washington), Genscan (Chris Burge,  
<http://genomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene  
<http://www.cbs.dtu.dk/netgene/cbsnetpgene.html>), searches of the  
 complete sequence against a peptide database and the Arabidopsis  
 EST database at TIGR (<http://www.tigr.org/tdb/at.html>).  
 Annotated genes are named to indicate the level of evidence for  
 their annotation. Genes with similarity to other proteins are named  
 after the database hits. Genes without significant peptide  
 similarity but with EST similarity are named as 'unknown' proteins.  
 Genes without protein or EST similarity, that are predicted by more  
 than two gene prediction programs over most of their length are

annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
 predicted by tRNAscan-SE (Sean Eddy,  
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are  
 identified by RepeatMasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of  
 genomic sequence that are not annotated as genes but have predicted  
 exons by GRAIL are annotated as misc features.  
 Location/Qualifiers  
 1. 89934  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="III"  
 /map="ml72"  
 /clone="T12J13"  
 1. 10209  
 /note="overlap with BAC clone F20H23  
 (AC009540:91202..101410)."  
 316..396  
 /note="exon predicted by xgrail, quality  
 marginal\_shadowexon"  
 complement(642..739)  
 /note="exon predicted by xgrail, quality excellent"  
 649..800  
 /note="exon predicted by xgrail, quality  
 marginal\_shadowexon"  
 complement(838..919)  
 /note="exon predicted by xgrail, quality marginal"  
 complement(join<924..1235,1334..1375,1465..1582,  
 1659..1753,1837..1944,2033..2106,2231..2303,2411..2458,  
 2535..2718,2815..2993,3155..3247,3378..3491,3720..3794,  
 3900..4163,4264..4377,4462..4518,4625..4781,4874..4938,  
 5205..5288,5367..5409,5510..5616,5744..5812,5995..6172,  
 6273..6466))  
 /gene="T12J13.1"  
 complement(<924..>6466)  
 /gene="T12J13.1"  
 /note="F20H23.26; similar to polynucleotide phosphorylase  
 GB:AA050039 [Pisum sativum], identical to putative  
 polynucleotide phosphorylase GB:AAF00646 [Arabidopsis  
 thaliana]"  
 complement(join(924..1235,1334..1375,1465..1582,  
 1659..1753,1837..1944,2033..2106,2231..2303,2411..2458,  
 2535..2718,2815..2993,3155..3247,3378..3491,3720..3794,  
 3900..4163,4264..4377,4462..4518,4625..4781,4874..4938,  
 5205..5288,5367..5409,5510..5616,5744..5812,5995..6172,  
 6273..6466))  
 /gene="T12J13.1"  
 /codon\_start=1  
 /product="putative polynucleotide phosphorylase"  
 /protein\_id="AAF03462.1"  
 /db\_xref="GI:6091752"  
 /translation="MIGSENISWEIKTRFNLPETRLATPHTESOEKLPRLQGTDKVS  
 LFLTSLSLKEEDVDVSHSGGKCLSLSSCGACKFSVRALVPPDPTDDADS  
 VGDGLAFPNHVSVKIPFCNBEILAVETGLMGQASAVTVDGETIVYISVCLADVPS  
 EPDFFPLVYHQHRSVAGVSGTSGGFRKGRKDKHEVLICRLIDRLPMPKGFYN  
 QTQLSWLSTQGLHAPDAVTSAGIAVASEVNPNAKAIAGVRVGLIGEEIVNPTV  
 KEMESQDLFDLGDGTDITLIEGYSNLFPEMLLQAVKVGQAVQATCAIEVLAKK  
 YGKPMIDAIRLPPPELYKHVKELAGEELTKALQIKSRKAISSLEKYLITLIE  
 KGVYDEVAFTGIEAQDLDLEDEVEVPEGVGDGVHPIPIPKIPLLFSEVD  
 KLVFKEVSKLLIRRIIVEGKSDGRTIDEIRPNISCGILPRANGSTLFTRGTOA  
 LAVALTIDKQMAQRIDNLEGSDEKFRYLQYTFPPSSVSGVGRIGAPSRREGCTLA  
 ERVETILSDDDFPYTIIVRSVFIESNSSSMASVCGGCLAQDAGVPPKVSAGTIA  
 MGVWTEFEGGSGSPILLSDITGADSGDMDFKAGNEDGVTAFOMDIKVGGTGLA  
 TMEKALIQAKGRHILAEAMKSPSTLSLKAPILILIMKHVPSKYSVSLIGSGGKK  
 KSLTIEESGVEAIDMQDDGVTKIMAIDVASLERAKITISITWVPSVGLIYRNCIKS  
 MAPYAFVETAPFREGELCHSELSEAWLAKPEDAYKVGDRIDVKLIEVNEKQLRVS  
 RALIPSEYDKDSQKQAGSDKSKSQKRYKNTSSKDRAAACAGSKVSGDELVLKK  
 KDVRATGGSSDKTWNSSSTNEESLVNGEATIS"  
 6302..6339  
 /rpt\_family="(CA)n"  
 complement(6956..7119)  
 /note="exon predicted by xgrail, quality marginal"  
 repeat\_region  
 misc\_feature







```

beta-synthase p-phosphateattachment site"
/codon_start=1
/transl_table=11
/product="cysK"
/protein_id="CAB06151.1"
/db_xref="GI:1781241"
/db_xref="SPTREMBL:P95230"
/translacion="MSIAEDITOLIGRTPLVRLRRVTDGAVADIVAKLEFFENPANSVK
DRICVAMLOAQAAGLIPDFIILEPTSGNTGIALAMVCARGYRCVIMPTMETSILR
RMLLATYAGELIITPCADGSGALAKAEALIKDORYFPQOFENPANFAIHRTVTAE
EVMRDYDGVIVAGVGTGTITGVAQVTKERKPFVAVPEAPASPPVSGGGPH
PIQIGAGFVPPVLDLDEIITVGNEDALNARRAREGLLVGISSGAATVAALQ
VARPENAGKLLIVVLPDGERYLSTPLFADVAD"
445..501
/misc_feature
/feature="cysK"
/feature="PS00901 Cysteine synthase/cystathionine
beta-synthase p-phosphate attachment site"
1285..1974
/feature="cysE"
/feature="cysE"
/feature="cysE"
/feature="Rv2335, (MTCY98.04), len: 229. cysE. Function:
probable serine acetyltransferase. len: 229 aa. FASTA
best: TR:G142363 G142363 ORF 7, (49.5% identity in 216 aa
overlap). TBPARE score is 0.894; contains PS00101
Bacterial hexapeptide repeat containing-transferases
signature"
/codon_start=1
/transl_table=11
/product="cysE"
/protein_id="CAB06152.1"
/db_xref="GI:1781242"
/db_xref="SPTREMBL:P95231"
/translacion="MLTARGDIAERDPAAPTALEVICPCVHVMGRLAHL
WQCARLLARAAETRLITGVDPHPCAVIGARVFIDHATGVIGETAEVGDVYIY
GVTGGSGMVGKRRGTVDRVITIGAKAVLGPIKIGEDSRIGANVAVPPPSAVV
VGVPGVIGQSPSGPGFWRPLDVLGASLDSLTUVRARLEALGGGQAAGVIRPPE
AGIWHGEDFSI"
1660..1746
/feature="cysE"
/feature="PS00101 Bacterial hexapeptide-repeat
containing-transferases signature"
2378..2382
/feature="gaggt, potential rbs upstream of Rv2336"
2390..3358
/feature="Rv2336"
2390..3358
/feature="Rv2336"
/feature="Rv2336, (MTCY98.05), len: 322 aa. Function:
unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2336"
/protein_id="CAB06153.1"
/db_xref="GI:1781243"
/db_xref="SPTREMBL:P95232"
/translacion="MDVPEQPALSSKSNRFTSQRTTGVGTVTVERLEPRLSPASR
HITEAFGTCEHVSFTREDDPDRVAVRVEIHCYAVVAAGHYVESALDELGLDLSN
AFELDKARGSTRETEVYLHVAPEPLSGQEGGLRIVGLISAVGSDLDLSAFKAAK
PSMGLAHQRLKYDAIEDLGHGVKEIALSVTADAPPTVSVSLIREVLRVHFTGEKL
IITFMFAYAKMNNFGRFMPQGEFFYAHNRNDRPSNDLLVPSIVPSFNSFLENI
SGVVYADGDGTARRRFTLCITMTDGLDDIFMPLTRQVLSGSIQDI"
complement(3422..4540)
/feature="Rv2337c"
/feature="Rv2337c"
/feature="Rv2337c, (MTCY98.06c), len: 372. Function:
unknownbut some similarity to another M. tuberculosis
protein, Q11025. FASTA best: Y071_MYCTU Q11025
hypothetical 78.2 kd protein SC102B10. (715 aa), fasta
scores, opt: 139, E(): 0.027, (25.6% identity in 367 aa
overlap) TBPARE score is 0.962"
/codon_start=1

```

```

/transl_table=11
/product="hypothetical protein Rv2337c"
/protein_id="CAB06154.1"
/db_xref="GI:1781244"
/db_xref="SPTREMBL:P95233"
/translacion="NRAGRNGPMTGLDPAEFLSLVEAAALAFSADNRREVLEHAGR
RVLWGSDOTWRSAPHEHRRIMSLVAIGAAVENVKLRAGRLGFETKVCWFPSDGNGLVA
EIVDRLPOTRYDPTLEGAIERRRTRNRVRFRGPLSPQCELGALSAAETGIDGQLHWF
DSPETRKOLLRLVRLAETERESEELHEELFSVAFEDIGMTASSDCLPGSLVEAM
MRPNFRGLHRRVRLRLRTVGMHHLGALRAAYLPCRLAPHVGLTTLSDLASGALTAG
AVFRIWMLRTLLGLAELQPFASAVLSLPAECWAPHVRAALVGMGNLLAPGHWPMV
FRIGHARASVTRMOSQSVAYCYAPAESGSDSESRA"
complement(4660..5616)
/feature="moew"
/feature="moew"
/feature="Rv2338c, (MTCY98.07c), len: 318 aa. Probable moew.
Some similarity to several molybdopterin biosynthesis
proteins: FASTA scores: sp|O27613|O27613 MOLYBDOPTERIN
BIOSYNTHESIS PROTEIN MO (251 aa) opt: 309 z-score: 387.4
E(): 4.7e-14; 30.7% identity in 254 aa overla; and
G1204377, (261 aa), fasta scores, opt: 271, E(): 3.3e-11,
(34.9% identity in 149 aa overlap)"
/codon_start=1
/transl_table=11
/product="moew"
/protein_id="CAB06155.1"
/db_xref="GI:1781245"
/db_xref="SPTREMBL:P95234"
/translacion="MRAGADAPDSGRVKESAPMSYDEAFCRNLGLISPTQOORLNSR
VAIAGSGVGGIDMVALARMGIGKFTIADPDVEIRNSNRQYGMARSTNGAKAEVMR
NIVHDINPEAEIRAFCEPIGKNAETLEGADVLVDGDAFEIDRLRLRYREATHRQSGI
YALGACGFSTAWVYFDPKGTDFRDYDSDAMTVDKVFVFIAGIAPSATRRSID
SLYVDINRRTGPSGLASGVAAEVLKILGHGRVYAAPYFHOFDAYRSIYVRK
BLRCNRRHRLQRKRLRLARYINRRSAGVPLGLRYHRTPSY"
6246..9134
/feature="mmpL9"
/feature="mmpL9"
/feature="mmpL9"
/feature="Rv2339, (MTCY98.08), len: 962 aa, Probable
membraneprotein with strong similarity to others in M.
leprae (YV34_MYCLE P34881; G699237 U1740AB and M.
tuberculosis (MTCY2069.34; MTCY48.08c; MTCY19G5.06), FASTA
best: YV34_MYCLE P54881 hypothetical 105.2 kd protein
u1740v. (959 aa) opt: 3799, E(): 0, (59.3% identity in 937
aa overlap) TBPARE scores 0.956"
/codon_start=1
/transl_table=11
/product="mmpL9"
/protein_id="CAB06156.1"
/db_xref="GI:1781246"
/db_xref="SPTREMBL:P95235"

```

## alignment\_scores:

Quality: 312.50 Length: 520  
Ratio: 1.250 Gaps: 24  
Percent Similarity: 48.077 Percent Identity: 25.962

## alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x MTCY98/rev ..  
Align seg 1/1 to reverse of: MTCY98 from: 1 to: 31225  
10 ValArgGluAlaProPheGlyTyrLysProGlySerLysGluSerIleG1 26  
||| ||| :||| ||| |||  
20176 GTATCGAAAGGCTATGGTCCGCTCCTCG.....GG 20139  
26 uasnLeuLysAspLysValGluAsnIleValTriPLeuLeuGluAsnA 43  
:||||| ||| :|||:||||| :||| :|||  
20138 TCATTGTACCAT ...ATCAGACACATGCTGCTGCTACAGAGAAC 20092  
43 rgSerPheAspAsnIleLeuGlyGlyValArgArg...GlnGlyLeuAsp 58  
|||||:|||||: :||| :||| :|||



```

FEATURES
  source      Location/Qualifiers
1..2262
/organism="Burkholderia pseudomallei"
/strain="spt1"
/db_xref="taxon:28450"
/clone="psn-1"
111..116
RBS
111..116
CDS
124..2226
/codon_start=1
/transl_table=11
/product="non-hemolytic phospholipase C"
/protein_id="AAFI7299.1"
/db_xref="GI:6572521"
/translation="MTNRRRDLRLAAGTAGAALQLFPVIREALIPAANRRGTGI
RDVHIVILMOENRSDHYFGKRGVDFGDPRLTLQNGKPVPHQVLLGPAELLPF
HPDASNGMOFLQDLPHGWODTHGANNKRGYRWIANKGTTMAYLERGDIPEHYOLA
DAFTICDAYHCSIPTSDPNRYNMWTCYVNCDCAGGVPVLCNEEAGYGRSTYPTLEQ
AGVSKYIYODICTGLDAAGSGWTONPYIGNYGDNALILFYENRQAPGSPLYDKART
GTVNSAGTLFDVLQDVNGTLFQVSWNICAPESYSEHPNPNYGNWIVYEQVLKALT
SNPDVSKTALFITYDENDGFHDVFPAPQSRDNGLSTVSTAGEIFPGDAHWAGP
YGLGRVPMVSPMTKGGWYCSQTFDHTSLQFTIARENDRIYVRAPNVTWRRRAVC
GDLTSAFNFSSPDGPNPQDPTSGYAPDRNRHPSYVFPVPAQSMFKQEAFLRAARA
LPYELFLVRIDQSTGKEKFLFAKTGRAGAAFOVTAGNRVDGPMWYTVEARLRLSDEW
STATLFLSYDLTVYGNPGFLCQPRFGTAAALGILNANPEVIYGYDVANGNTLRLNRG
RATVRLTVANAYCKSPRYELKPGQRIINDYNDLRDLSHSWYDLTVSDGAPNGFLRFA
GHVETGRPSDPLATV"
BASE COUNT      388 a 794 c 752 g 328 t
ORIGIN

alignment_scores:
  Quality: 309.00      Length: 430
  Ratio: 1.405        Gaps: 23
  Percent Similarity: 51.163      Percent Identity: 27.674

alignment_block:
  US-09-426-072-2_COPY_20_464 x AFI07252 ..
  Align seg 1/1 to: AFI07252 from: 1 to: 2262
10 ValArgGluAla.....ProPheGlyTyrLysProGlySerLysG1 23
   :|||||
208 ATCCGCGAAGCGCTCGGATTCCCGCGAAGCGCGCGCACG..... 252
23 usertleGluAsnLeuLysAspLysValGluAsnIleValTrpLeuIle 40
   :|||||
253 .....ATCCGCGAC...GTCGAGCACATCGTGATCTGTATGC 286
40 euGluAsnArgSerPheAspAsnIleLeuGly.....GlyValArg 53
   :|||||
287 AGMAAACCGCTGTTTCGACCATTTCTCGCAGCTGCGCGCGCGTGC 336
54 ArgGlnGlyLeuAspAsnProle.....AsnAsnGlyProPheCy 67
   :|||||
337 GCCTTCGCGCATCGCGCGCGCTCACGCTGCAGAACGCGCGGTGT 386
67 sasTyrLysAsnAlaSerAspProSerSerGlyLysTyrCysThrGln 84
   :|||||
387 CCATCAGCGGTGCTCTCGCGCGCGC..... 414
84 laLysAspTyrAspSerValPheAsnAspProAspHisSerValThrGly 100
   :|||||
415 .....GAGCTGCTGCCGTTTCCATCCGAGCAGCATCGAATCTCGC 453
101 AsnAsnLeuGluPhe.....TyrGlyThrTyrThrProAs 112
   :|||||
454 .....ATGCAAGTCTCTGCAAGACCTGCCGACGCG...TGGCAGGAC 494
112 nAsnGlyAlaIleAserGlyLysValValAlaAspGlnSerGlyPhe 129
   :|||||
495 GCACGCGCATGGAACAAAGGCGCG..... 519
129 euAsnAlaGlnLeuAsnAspTyrProLysLeuAlaProGluGluAlaThr 145

```

```

520 .....TACGACCGCTGGATCCGGAACAGGCGACG 549
146 ArgGlnValMetGlyTyrThrGluGluValProThrLeuValas 162
   :|||||
550 ...ACGACGATGGCTACCTCGAAGCGGGGACATCCCGTTCCACTACCA 596
162 pLeuValAspGluPheThrPheAsnSerTrpPheSerCysValPro 179
   :|||||
597 GCTCCCGATGCTTCCAGATCTCGGAGCGGTACCACTGCTCGATCCGA 646
179 lyProThrAsnProAsnArgLeuCysAlaLeuAlaGlyThrAlaGly 195
   :|||||
647 GCTCGACCGCGCGGACCGGTACTACATGTGCGAGCGGTACGTCGGCAAC 696
196 HlslGlyLysAsnAspAspPheLeu.....AsnTyrG1 207
   :|||||
697 GACGCGCGCGCGCGCGCGGTGCTCGGCAACGAGAGCGCGGTACGG 746
207 yleSerSerLysSerIlePheGluAlaAlaAsnGluLysGlyValSerT 224
   :|||||
747 CTGGAGCAGCTATCCG.....GAGAGCTCGAGAGCGCGCGGTGCTGT 790
224 rPLeuAsnTyrAsp.....GlyThrAsn..... 231
   :|||||
791 GGAAGATCTATCAGGACATCGCGACGCGCGCTCGACGCGCGCGGTG 840
232 .....GlyGluPheGluProAspSerLeuPh 240
   :|||||
841 GGTGGACGAGAACCCGTACATCGGCAACTACGCGCGACATGCGCTCT 890
240 ePhe.....ThrTyrVala 245
   :|||||
891 CTACTTCAACCATGACCGCAACGCGCGCGCGCGCTGTACGACA 940
245 snGlnThrSerArgSerAsnValValProValGluAsnPhe..... 259
   :|||||
941 AGGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 990
260 .....GlnAspAlaTyrLeuGlyValLeuProLysSerTyrIleAs 274
   :|||||
991 TTGCAGCAGGACGTGAAGAAGCGCGCGCGCGCGCGCGCGCGCGGTG 1038
274 nProSerCysCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnV 291
   :|||||
1039 .....TGC CGCGCGGAAAGCGTATTTCGAGCACCCGCAACTGCGCG 1078
291 alSerTyrGlyValPheValLysGlnIleTyrAspAlaIleArgGln 307
   :|||||
1079 CGAACTACGCGCGCTGTGTCAGTCGAGCAGGTGCTGAAGCGCGCTGAC 1128
308 GlyProGln...TrpAspLysThrLeuPheIleThrTyrAspGluTh 323
   :|||||
1129 AATCCGCGCGGTGGAGCAAAACCGCGCTCTTCATCACGTACGACGAGA 1178
323 rGlyGlyPheTyrAspHisValProProLeuAla.....ValArgp 338
   :|||||
1179 CGAGCGCTTCTTCGATCAGCTCGCGCGCGCTTCGCGCGCAATCGCG 1228
338 roAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyr... 353
   :|||||
1229 ACAACGGCTGTGACGCGTGTGACCGCG.....GGCGAGATCTTCCCG 1272
354 .....ThrLeuHisPheAsp.....ArgLeuGlyGlyArgMetPr 365
   :|||||
1273 GGCAGCGCGCGCACATGCTGGCGCGGTACGGGCTCGGGCGCGCGTCC 1322
365 othrTrpValIleSerProTyrSerLysLysGlyTyrIle 378
   :|||||
1323 GATGCTGCTGATATCGCGGTGGACGAAGCGCGCGCTGGGTC 1362
seq_name: gb_ba2.MBOY18606

```



/gene="RvD2-Rv1758"  
 /note="corresponds to M. tuberculosis complete genome  
 region 1989055..1989337, AccNo AL123456"  
 BASE COUNT 1420 a 2422 c 2599 g 1496 t  
 ORIGIN

## alignment\_scores:

Quality: 302.50 Length: 423  
 Ratio: 1.505 Gaps: 18  
 Percent Similarity: 47.518 Percent Identity: 27.896

## alignment\_block:

US-09-426-072-2\_copy\_20\_464 x MBOY18606/rev ..

Align seg 1/1 to reverse of: MBOY18606 from: 1 to: 7937

```

10 ValArgGluAlaProPheGlyTyrLysProGlySerLysGluSerIleG1 26
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1451 GTGATTGAAGGCGCTACGGCGGGGCCCTGCTCC.....GG 1414

26 uAsnLeuLysAspLysValGluAsnIleValTrpLeuIleLeuGluAsnA 43
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1413 CCATTTCAGCCGAC...ATCGAACACTTGTGTTCTTCATGCAGGAGAAC 1367

43 rGserPheAspAsnIleLeuGlyGlyValArgGlnGlyLeuAspAsn 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1366 GCTCATTCATCATTACTACTTCGGAACATC.....TCCGGCACCGATGGA 1323

60 ProIleAsnGlnGlyProPheCysAsnTyrLysAsnAlaSerAspProSe 76
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1322 TTCAACACGGGTATCGCGCTCTTCCACAAAGGGTGGAAACCG..... 1278

76 rSerGlyLysTyrCysThrGlnAlaLysAspTyrAspSerVal..... 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1277 .....ATGACGACGGCTCTTGACGCCCGCGCTTACGATGTC 1241

91 .....Phe 91

1240 CCTACCGCTTCGACACCGACCGCGCGCTTCCTCGACGGCGGTGTC 1191

92 AsnAspProAspHisSerValThrGlyAsnAsnLeuGluPheTyrGlyTh 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1190 ACAGACCTGACCATAGC.....TGGGTGGCGAT 1162

108 rTyrThrProAsnAsnGlyAlaIleAlaSerGlyLysValAlaAspG 125
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1161 GCATGAGTCGTGGAATGTTGGGTC.....A 1136

125 InSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeuAlaPro 141
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1135 ACAGCACTGGCTACCGACCGCAG.....CCCAAGACTCGGTGTC 1098

142 GluGluAlaThrArgGlnValMetGlyTyrThrGluGluGluValPr 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1097 GCGGCCACACATCCCGCGGTGATGGGTACTACACGCGCAAGACATTCC 1048

158 oThrLeuValAspLeuValAspGluPheThrPheAsnSerTrpPheS 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1047 GATCCACTATCTGTTGGCGAGCGCTTACCGCTCTCGACCGCTACTTTT 998

175 eCysValProGlyProThrAsnProAsnArgLeuCysAlaLeuAlaGly 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
997 GCTCGGTATTGGGTCCACACCTGCGCCCAACCGCCTTTATTGGCTAGCGCC 948

192 ThrAlaAlaGlyHisGlyLysAsnAspAspPheLeu..... 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
947 ACATCGACCCCGACGCGCAAAATGGTGGCGCGGACCTCGAGATCCAC 898

205 .....AsnTyrGlyIleSerSerLysSerIleP 214
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
897 TTTCACACGGTGGGAGATTTCGGTTGGCGCATCATCGCGGAGAACCTTA 848

```

```

214 heGluAlaAlaAsnGluLysGlyValSerTrpLeuAsnTyrAspGlyThr 230
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
847 GCGATGCC.....GCGTCACGTGAAGGTGAT.....CGC 816

231 AsnGlyGluPheGluProAspSerLeuPhePheThrTyrValAsnGlnThr 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
815 AACAAAGACACTCGGGCCCATCTCCTCGGTCTTACTTACGGCTCGCTGT 766

247 rSer.....ArgSerA 251
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 GACGCTTTTCAACAGCTACGCCGATCCAGGTGATCTTGTCCGCTTTG 716

251 snValValPro.....ValGluAsnPhePheGlnAspAlaTyrLeuGly 265
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
715 GCGTGGCACCAAGCTATCCCGCAGCTTCGCGCGCGCGCTTTACCCAAT 666

266 ValLeuProLysPheSerTyrIleAsnProSerCysGlyThrAsnTh 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
665 AGACTGCCGCGGGTCTCCTGGGTGATCCCAATGTT.....CT 628

282 rAsnSerMethHisProThrGlyAsnValSerTyrGlyGluValPheVal 299
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
627 CGAATCCGAACATCTCGCGTTCACGCGCGCGCGCTTTTCGCAATCG 578

299 ysGlnIleTyrAspAlaIleArgGlnGlyProGln...TrpAspLysThr 314
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
577 TCAACATCTTAAGAAATTTGCTTCCAAATCTCGCGTGTGGGAAAGACG 528

315 LeuLeuPheIleThrTyrAspGluThrGlyPheTyrAspHisValPr 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 GCGCTGATCGTCAGCTACGACGAAACGCGGCTTTTTCGACACGTTGT 478

331 oProPLeuAlaValArgProAsnLeuThrTyrThrGluThrAlaL 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
477 TCTCTCTACCGCG.....CCGCGCGGGA 455

348 ysAsnGlyGlnLysTyrThrLeu...HisPheAspArg..... 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 CTCCGCGGGAATATGTCACGCTGCTGACATCGATCAGGTCCGGCTCC 405

360 .....LeuGlyGlyArgMetProThrTrpVa 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 GCGGAATACGCGGCGCGATCGGTGGGCTTTCGCGTTCCTCTGTTGT 355

368 lIleSerProTyrSerLys 374
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 CATTTTCGCGTACAGCGGT 336

```

seq\_name: gb\_pl1-AC005168

## seq\_documentation\_block:

LOCUS AC005168 91893 bp DNA PLN 05-APR-2000  
 DEFINITION Arabidopsis thaliana chromosome II section 150 of 255 of the  
 complete sequence. Sequence from clones F18A8, F13C20, T20P8.  
 ACCESSION AC005168 AE002093  
 VERSION AC005168.2 GI:6598444  
 KEYWORDS HTG;  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids II;  
 Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 91893)

AUTHORS  
 Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-L., Town C.D.,  
 Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E.,  
 Feldblum T.V., Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M.,  
 Koo H., Moffat K.S., Cronin L.A., Shen M., VanAken S.E., Unavam L.,  
 Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
 Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M. and  
 Venter J.C.

TITLE  
 Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana









```

353 yThrLeuHisPheAspArg.....LeuGlyClyArgMet 364
      :::::  :::::  |||||  |||:::
1456 ....ATCGAGGATATCCACGGGTCGCCATCGCGCTCGGCACCGGTC 1500
      :::::  :::::  |||||  |||:::
365 ProThrTrpValIleSerProTyrSerIlysGlyTyrIleGluGlnTy 381
      :::::  :::::  |||||  |||:::
1501 CCGTGATCGGCATCTCGCCTGTCGAAGGGCGCAAGTC..... 1542
      :::::  :::::  |||||  |||:::
381 rGlyThrAspProValThrGlyLysProAlaProTyrSerAlaThrSerV 398
      :::::  :::::  |||||  |||:::
1543 .....AGCGCGAAGTGTTCGACCATACCTCGG 1570
      :::::  :::::  |||||  |||:::
398 aLeuLysThrLeu..... 402
      |||||  |||
1571 TCGTCGGCTTCTCGAACCGCTTCGGCCTCGTCGAGAAACATCTCG 1620
      |||||  |||
403 .....GlyTyrLeuTrpAspIleGluAspPhe... 411
      |||||  |||
1621 CCTCGCGGCTCGCTCGCGGCACTGACTTCCTCGACTTCCA 1670
      |||||  |||
412 .....ThrProArgValAlaHisSerProSerP 421
      :::::  :::::  |||||  |||:::
1671 GGGCGCAGCGGATACCCAGGTGCGCCCGCCACTGACCAACGTCGCGCAGA 1720
      :::::  :::::  |||||  |||:::
421 heAspHisLeuIleGlyThrThrLeuArgGluAspAla 433
      |||||  |||
1721 GCGAC.....GCGCGCAAGGAGGACGCC 1743

```

seq\_name: gb\_bal:AE004519

```

seq_documentation_block:
LOCUS      AE004519      11622 bp      DNA      BCT      30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 80 of 529 of the complete
            genome.
ACCESSION  AE004519 AE004091
VERSION    AE004519.1 GI:9946736
KEYWORDS
SOURCE
ORGANISM   Pseudomonas aeruginosa.
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE  1 (bases 1 to 11622)
            Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
            Hickey,M.J., Brinkman,F.S., Huftnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z., Paulsen,I.T., Paulsen,I.T.,
            Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
            Nature 406 (6799), 959-964 (2000)
            20437337
            2 (bases 1 to 11622)
            Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
            Hickey,M.J., Brinkman,F.S.L., Huftnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
            Direct Submission
            Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
            Location/Qualifiers
            1. .11622
               /organism="Pseudomonas aeruginosa"
               /strain="PA01"
               /db_xref="taxon:287"
            126. .1085
               /gene="PA0841"
            126. .1085
               /gene="PA0841"

```

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04230.1"
/db_xref="GI:9946737"
/transl_table="MRIGLVVDATCDLPPFEFLAAHCRVMPGIRLGEORLDRDRPD
TTLSFYRDHLPKVKGSGKSOPLSAEQLQWDELDTVDYVCLTVDSLRSPFEHA
TOASFGLOOHYHERRHAGIAGPTFLVLLDSISYFAGIACLAAGTRLLAOGSHPNAL
RTIRLEQCTCTYLVADDLGLRRGFGKQDSRGFVDRVGAALGIGSLDDVKPFVS
LADGEDRPVALSPFSAEKLFGYALARVQAGELLAPOLGLSYAGDPTALRDLPGFA
ALENACREDIHLHGLMSPGTCGGINLAGAGMSLGTATPKAFA"
/complement(1089..2309)
/gene="PA0842"
/complement(1089..2309)
/gene="PA0842"
/codon_start=1
/transl_table=11
/product="probable glycosyl transferase"
/protein_id="AAG04231.1"
/db_xref="GI:9946738"
/transl_table="MSRPLSIALISSETYPPEVNGVANTLGRHAGLQOLGHRVQVVRP
RQPGDGRSDAELVITRCWPLPGYGLQWSSLLHLKRCMKRQRPDVLVLTAEPL
GFSALRAARLGIPIAVSGFHTNFOYSEHYGCPILTRLVTVGLWHEHNTQMTLVPSG
SORMELORRGFERLNLISRGVDSOLPHSPRDELRBRWGLGEOIDVAVLHVGELAAEK
NLGLLGGTFPACAHLPOLKRLVLVGDGPERKHRLEROLPALFCGQVGETLAAHYA
SGDLFLPFSLETFGNVLEAAGLAVAFDAQAGQIRHGHINGVLAAPGDKSGFV
EVASWLLDQOERLRRVLRNARHQAHSRQWDSIVFHFQNYLEDACASPLVQPARSKPER
PPAEDGGALPRSRE"
/complement(2397..3020)
/gene="plcR"
/gene="PA0843"
/complement(2397..3020)
/gene="plcR"
/codon_start=1
/transl_table=11
/product="phospholipase accessory protein PlcR precursor"
/protein_id="AAG04232.1"
/db_xref="GI:9946739"
/transl_table="MRPTLTWTLLALLCGTAIGAVLLFYSEPAVPASPPOATP
AAKPSIPSAPEMTATADNLEOQLGFCRAGOMSEIRKOAEGLEOLKREVAV
GADPRQTFEIOHLPYVEADARRREALDEIWMALKONASYQQOAPTGCEEEQLREY
AQESDKVIAEVLASVDEEQRRHAALDERLAKLRKQIFGENPRLIQR"
/complement(3038..5230)
/gene="plcH"
/gene="PA0844"
/complement(3038..5230)
/gene="plcH"
/codon_start=1
/transl_table=11
/product="hemolytic phospholipase C precursor"
/protein_id="AAG04233.1"
/db_xref="GI:9946740"
/transl_table="MTENMKERRFTFLKHGAQAATLAGLSGLFPETLRRALAVEPDIR
TGTITQDVHVILMQENRSFOHYEGLNGVRGNDPRAKQRQDGKPVWQNTKYFEFS
YHMTKVTSAQWYSSQNHWSAFIAINWGRNDKMAVQYPEAMGYFKRGDIPYVAL
ADAPTLCIAYHQSMGPTNPNRLYHMSGRAAPSGDGKDVHIGNDMGDTIGASGTVDW
NTPRSDLSQAMARLVGYOEGYRSSLWLYLVDAWKYRLOEQNNYDCNALAFNRFX
TAPRDLQWORAAMARLVGYOEGYRSSLWLYLVDAWKYRLOEQNNYDCNALAFNRFX
RVLEALNENPWAHTVELLNYDCSGDFYDHASAPVPKWCYGLSTVSTAGEIEVSS
GLITGLGHRVPLIATSPWSKGVSAEVDHTSVLRFLERFVGVEENISPMRAVCG
DLTSLDFQAGGTQVAPDLTNVPSDARKEDAYWQOYFRPSKYNSYSEKSLPGQEK
GORETFLAVYQLHATLDAIAAGKRLTLNGLDMSLPGNPGQHSAAVQVQPRNLHAA
RFYTVTSYVVOEGSEIGRTLNDELDDLDANGRYAFEHVHGPFPRFPHGNLHAA
QMARPEVSYVQNRGNLQNLINRLNGRPSVTVTPNPATYREGSRRYELPEQNAISEV
WLLRSQGWYDLSTASNTAANYLRLRLAGHVETGKPSRSDPLLDIAAT"
/complement(5572..7584)
/gene="PA0845"
/complement(5572..7584)
/gene="PA0845"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04234.1"

```

Align seg 1/1 to reverse of: AE004519 from: 1 to: 11622

```
32 ValGluAsn1leValTrpLeuileLeuGluAsnArgSerPheAspAsn11 48
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
5080 GTCCAGCACGCTGTCTATCTGATCGACGAGAACCGCAGCTTCGATCACTA 5031
48 eleuGly.....GlyValArgArgGlnGlnGlyLeuAspAsnPro.... 60
:::||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
5030 CTTGGCCACCTCAACGGCGTCCG.....GGCTTCAACGACCTCGCG 4987

61 .....IleAsnAsnGlyPro.....PheCysAsnTyrIlyAsn 71
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4986 CGCTCAAGCCCGCAGGACGCGGCTGCTGCTACAGAACATPACAAGTAC 4937
72 AlaSerAspPro.....SerSerIlyLysTyrCy 81
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4936 GAGTTCCTCCCTCATCTGCGGATACCAAGGTCAACAGCGCCCGCTGGT 4887

81 sThrGlnAlaLysAspTyrAspSerValPheAsnAspProAspHisServ 98
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4886 GAGTTCGACGAACCATGAATGGTCGGGCTTC.....CACGCGA 4849
98 alThrGlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAsnAsnGly 114
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4848 TCTGGS.....AACCAGGC 4835

115 AlaIleAlaSerGlyLysValValAlaAsnGlnSerGlyPheLeuAsnAl 131
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4834 .....CGCAACGACAAGTGGATGGCGGTGCAG..... 4808
131 acIleLeuAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnV 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4807 .....TACCCC.....GAAG 4798

148 alMetGlyTyrTyrThrGluGluValProThrIleuValAspLeuVal 164
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4797 CCATGGGCTACTCAAGCGCGGACATCTCTACTACTACGCCCTGGCC 4748

165 AspGluPheThrThrPheAsnSerTrpPheSerCysValProGlyProTh 181
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4747 CACGCCCTTCCACCTGTGCGAGGCTACACCATCTGATGATGGCGCCGAC 4698
181 rAsnProAsnArgLeuCysAlaLeuAlaGlyThrAlaAlaGlyHisGlyL 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4697 CAATCCGAACCGCTCTACCATGACGAGCGCGCGCCGCCACGCGCG 4648

198 ysAsnAspAspAspPheLeuAsnTyrGlyIleSerSerIysSerIlePhe 214
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4647 ACGGCAAGACGTACATCTGGCAACGACATGGGCGGCGGACCATCGCG 4598
215 .....GluAlaAlaAsnGluLy 220
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4597 GCCAGCGGCACGCTGGACTGGACCACTATCCCGAGCGCTAAGCGCGCG 4548

220 scIlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheGluProA 237
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4547 TGGCGTGGACTCGGGGTCTACCA.....GAAGCGGCTACCGCTCCT 4504
237 spSerLeuPhePheThrTyrValAsnGlnThrSerArgSerAsnValVal 253
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4503 CGTCACTCTGGTACCTGTACGTTCGACGCTACTGGAAATACGGGTCCAG 4454
254 ProValGluAsnPhe..... 258
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4453 GAGCAGACAACTACGACTGCACGCCCTCGCTGGTTCAGGAATCTCAA 4404

259 .....PheGlnAspAlaTyrLeu..... 264
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4403 GAACGCTCCGCGATTCGGACCTCTCTGCACGCGCGCATGCTCGCGCGCG 4354
265 .....GlyValLeuProLysPhe 270
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
alignment_scores:
  Quality: 287.50
  Ratio: 1.250
  Percent Similarity: 46.371
  Length: 496
  Gaps: 22
  Percent Identity: 23.790

alignment_block:
  US-09-426-072-2_COPY_20_464_x_AE004519/rev_1
```

```

4353 GCGTCGACCACTGCCAAGGATGTCAGAGAACACGCTGCCGAGGTG 4304
271 SerTyrIleAsnProSerCysGlyThrAsnThrAsnSerMetHisPr 287
||||:||||| ||| :||| |||||
4303 TCGTGGATGTCGCGGCCCTACTGCTACTGCGAG.....CATCC 4266
287 oThrGlyAsnValSerTyrGlyGluValPheValLysGlnIleTyrAsp 304
|||:||||:||||| |||:||||:|||||
4265 CTGTGGGGCCCTCGTCTCGGAGTACTAGTACCGCGGTGCTCGAGG 4216
304 IalIleArgGlnGlyProGln...TrpAspLysThrLeuLeuPheIleThr 319
||||:||||:||||| |||:||||:|||||
4215 CGCTACACAGCACCCCGAGGCTCGGCCAGGACCGTGTTCATCCTCAAC 4166
320 TyrAspGluThrGlyGlyPheTyrAspHisValProProLeuAlaVal 336
||||:||||| ||||| ||||| |||||
4165 TAGCAGAGGCGGAGCGCTTCTATGACCGCCGCGCCGCGTCCGCC 4116
336 IargProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysT 353
||||:||||:||||| |||:||||:|||||
4115 GTGGAAGGATGGCTGGCGCTTTCACGCGTACGACCGCGCGAG.... 4070
353 yThrLeuHisPheAspArg.....LeuGlyGlyArgMet 364
||||:||||:||||| ||||| |||||
4069 .....ATCAGGATATCCAGCGGCTGCCATCGGCTCGCCACCGGTC 4025
365 ProThrTrpValIleSerProTyrSerLysLysGlyTyrIleGluGlnTy 381
||||:||||:||||| ||||| |||||
4024 CGCTGATCGGCATCTCGCCCTGCTGCGAAGGGCGGCAAGGTC..... 3983
381 rGlyThrAspProValThrGlyLysProAlaProTyrSerAlaThrSerV 398
||||:||||:||||| ||||| |||||
3982 .....AGCGCGAAGTGTTCGACCATACCTCGG 3955
398 aLeuLysThrLeu..... 402
||||:||||:|||||
3954 TCGTCGCTTCCGACGCGCTTCGCGCTCGCGAGGAAACATCTCG 3905
403 .....GlyTyrLeuTrpAspIleGluAspPhe.. 411
||| ||| |||:|||||
3904 CCCTGGCGGCTGCGCTCTGCGGCGACCTGACTCGCTGTCGACTCCA 3855
412 .....ThrProArgValAlaHisSerProSerP 421
||||:||||:||||| ||||| |||||
3854 GGGCGGCGGATACCGAGTGCCTCCCGACCTGACCAACGTCGCGCAGA 3805
421 heAspHisLeuIleGlyThrThrLeuArgGluAspAla 433
||| |||:||||:|||||
3804 GCGAC.....GCGCGCAAGGAGGACGCC 3782
seq_name: gb_ba2.MSGMTP40A

seq_documentation_block: 3013 bp DNA circular BCT 16-JUN-1994
LOCUS MSGMTP40A
DEFINITION Mycobacterium tuberculosis phospholipase C gene and MTP40 protein,
complete cds.
ACCESSION L11868
VERSION L11868.1 GI:399585
KEYWORDS MTP40 protein; phospholipase C.
SOURCE Mycobacterium tuberculosis (strain H37Rv) (library: sub-library of
BAW HI M. tuberculosis DNA) DNA.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
1 (bases 1 to 3013)
AUTHORS Parra,C.A., Londono,L.P., del Portillo,P. and Patarroyo,M.E.
TITLE Isolation, characterization, and molecular cloning of a specific
Mycobacterium tuberculosis antigen gene: identification of a
species-specific sequence
JOURNAL Infect. Immun. 59, 3411-3417 (1991)
MEDLINE 91372944
REFERENCE 2 (bases 1 to 3013)

```

```

AUTHORS del Portillo,P., Murillo,L.A. and Patarroyo,M.E.
TITLE Amplification of a species-specific DNA fragment of Mycobacterium
tuberculosis and its possible use in diagnosis
JOURNAL J. Clin. Microbiol. 29, 2163-2168 (1991)
MEDLINE 92042637
REFERENCE 3 (bases 1 to 3013)
AUTHORS Leao,S.C., Rocha,C.L., Parra,C.A. and Patarroyo,M.E.
TITLE Nucleotide sequencing of a 3.0 kbp fragment of Mycobacterium
tuberculosis DNA
JOURNAL unpublished (1993)
FEATURES
source Location/Qualifiers
1..3013
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/tissue_lib="sub-library of BAW HI M. tuberculosis DNA"
462..2000
/citation=[3]
/codon_start=1
/transl_table=11
/label=M-TBC-EXCLUSIVE
/product="phospholipase C"
/protein_id="AAB59164.1"
/db_xref="GI:399586"
/transl_table="MSRREFLKTGAGAAFLMDHAAPIEKAYGAGCPGHLTDIE
HYLLMOENRSPDHYFTGLSTNGFNAASPAFOQMGWNPMTQALDPAGVTIPFRDIT
RGPFDEGVNDPEHQWGMHLMWNGGANDWLPQAATTRAGPYVPLTMGYTTRDIT
IHYLLADTFTICDGYHCSLLTGLPNRLYLWLSANDPAGTGGQPLVPGFLPQQFS
WRIMLEADGYSWKVYQNGKGLRFINTPISNGLVQAFQAAADPRSNLARYGIAPT
YPGDFAADVRRANLPKVSMLPNILQSEHPALVALGASVMTALRILLNPAVMEKT
ALIVSYDENGDFDHYTPAPPGTGEFTVPIVDIVDVGSGGIRGLGIGFRVPCIV
ISPYSRGLMVSPTFDHTSOLKLLIRARFVVPNMTANRDGVGDMTSAFNATPPNS
TRPNLSHPLLGALPKLPQICPNVILGTTDGLPSIPRYVPYQVMPTQETTPVRGTGS
GLCS"
complement(1417)
/citation=[1]
/citation=[2]
/replace="a"
complement(1478..1915)
/citation=[1]
/citation=[2]
/codon_start=1
/transl_table=11
/label=M-TBC-EXCLUSIVE
/product="MTP40 PROTEIN"
/protein_id="AAB59165.1"
/db_xref="GI:399587"
/transl_table="MIGNAPSVPVNPNTTLMHCGSGFASPSNGLKGLVDFGCVAKLN
AEVMSPTTSPRQAVMLGCTPNRARNINCEVMSVSETISCPRLYGEHTMQGTRKPR
PSGPRMPDPGCTASMLGVTNPSGPGVGGVGTWSKKRPSS"
complement(1530..1532)
/citation=[1]
/citation=[2]
/replace="cc"
complement(1940..1941)
/citation=[1]
/citation=[2]
/replace="g"
complement(1980..1982)
/citation=[1]
/citation=[2]
/replace="cc"
BASE COUNT 585 a 987 c 883 g 558 t
ORIGIN

alignment_scores:
Quality: 276.50 Length: 402
Ratio: 1.487 Gaps: 16
Percent Similarity: 46.269 Percent Identity: 28.358
alignment_block:
US-09-426-072-2_COPY_20_464 x MSGMTP40A

```

Align seg 1/1 to: MSGWTP40A from: 1 to: 3013

```

32 ValGluAsnIleValTrpLeuIleLeuGluAsnArgSerPheAspAsnI 48
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
588 ATGAGCATATCGTTCGTGATGCAGGAGAACCGGTATTCGACCACTA 637
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
48 LeuGlyGlyValArgGlnGlyLeuAspAsnProIleAsnAsnGlyP 65
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
638 TTTCGGAACGCTTCCAGC..... 656

65 roPheCysAsnTrpIleAsnAlaSerAspProSerSerGlyLysTyr... 80
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
657 .....ACCAATGGTTCACGCCCGCTCCGGCATTCACAAACAATGGGT 701
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
81 .....CysThrGlnAlaLysAspTyrAspSerVal..... 90
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
702 TGGAAACCCATGACGACGAGGCTTGGACCGCGCGGGTCCACATTCGGTT 751
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
91 .....PheAsnA 93
|||||
752 CCGCTTGGACACCCCGAGGCCCTTCCTGACGCGCGAGTGCCTCAACG 801
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
93 spProAspHisSerValThrGlyAsnAsnLeuGluPheTyrGlyThrTyr 109
||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
802 ACCCGAGCACAGTGGGTGGGATGCACCTGGCTGG..... 839
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
110 ThrProAsnGlyAlaIleAlaSerGlyLysValValAlaAspGlnSe 126
||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
840 .....AACGGTGGTGC..... 851
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
126 rGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeuAlaProGluG 143
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
852 .....AACGAC.....AACTGGCTCGCGCGC 873
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
143 luAlaThrArgGln.....ValMetGlyTyrTyr 152
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 AGCGACACCGCGCAGGACCATATGTCCTTTGACCATGGGTACTAC 923
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
153 ThrGluGluValProThrLeuValAspLeuValAspGluPheThrTh 169
||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
924 ACGCGCAAGATCCCGATCCACTATCTGTCGCGCACACGTTCCACCAT 973
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
169 rPheAsnSerTrpPheSerCysValProGlyProThrAsnProAsnArgL 186
||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
974 CTGCGACGCTACCATGCTGCTGCTGACGGCACCTGCCCAACCGGC 1023
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
186 euCysAlaLeu.....AlaGlyThrAlaAlaGlyHis 196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1024 TCTACTGGTTAGCCCAACATCGACCCCGCGGACCGACGGGGACCC 1073
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
197 GlyLysAsnAspAspPheLeuAsnTrpGlyIleSerSerLysSerI 213
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1074 CAATTGGTAGACCGGGCTTCCTCGCGCTGCAGCAATTCAGTTGGCGCAT 1123
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
213 ePhe...GluAlaAlaAsnGlyLysGlyValSerTrpLeuAsnTrpAspG 229
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1124 CATGCCGGAACCTCGAAGATCCCGGGTCACTGCTGGAAGGTGACCA 1173
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
229 lyThrAsn...GlyLupPheLupProAspSerLeuPheThrTyrVal 244
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1174 ACAAGGCGCTCGGGCGATTTCATCAACAGCGCCATCAGCAATAACGGGCTG 1223
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
245 AsnGlnThrSerArgSerAsnValValProValGluAsn..... 257
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1224 GTGAGGCTTCGCCAGGAGCGTATCGAGAGTGCAGACTTGGCCCGCTA 1273
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
258 .....PhePheGlnAspAlaTyrLeuG 265
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1274 CGGTATCGCCCGACCTACCTGGGACTTCGCTCGCGAGCTCAGGGCCA 1323
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
265 lyValLeuProLysPheSerTyrIleAsnProSerCysCysGlyThrAsn 281
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

```

```

1324 ACCGGTACCAAGGTCTCTCTGGTTAGTTCACCAATC..... 1361
282 ThrAsnSerMetHisProThrGlyAsnValSerTyrGlyCluValPheVa 298
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1362 CTCAGTCCGACACACCCCGCTGCGGTAGCGCTTGGCGGGTGTTCAT 1411
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
298 LysGlnIleTyrAspAlaIleArgGlnGlyProGln...TrpAspLysT 314
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1412 GGTGACCGCGTGGCGATCTTGTCTCCAATCCCGGGTGTGGAAAGA 1461
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
314 hrLeuLeuPheIleThrThrAspGluThrGlyGlyPheTyrAspHisVal 330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1462 CCCACTATGTCAGCTATGAGAGAACGGCGCTTCTTCGACCACTGC 1511
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
331 ProProLeuAla.....ValArgProAs 339
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1512 ACGCCCGCCACGGCAGCGCGGACACCGCGGCAATTCGTCACGGTGCC 1561
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
339 pAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuH 356
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1562 CAACATCCAC...GCAGTACCGGGTCCGGTGGCAATTCGTTGGTCCGCTC 1607
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
356 isPheAspArgLeuGlyGlyArgMetProThrTrpValIleSerProTyr 372
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1608 .....GGTCTGGGTTTCGGTTCCTGCTGCATTGTTCATTTCCGCGTAC 1649
:||||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
373 SerLys 374
|||||
1650 AGCGCG 1655

seq_name: gb_ba2.MTU49511

seq_documentation_block: 4094 bp DNA BCT 05-JUN-1998
LOGUS MTU49511
DEFINITION Mycobacterium tuberculosis phospholipase C (mpcA) and phospholipase
C (mpcB) genes, complete cds.
ACCESSION U49511
VERSION U49511.1 GI:1399492
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
1 (bases 1 to 4094)
AUTHORS Johansen, K. A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-1996) Kristine A. Johansen, Dept. of
Microbiology, Univ. of Colo. Health Sciences Center, 4200 East
Ninth Ave., Denver, CO 80262, USA
FEATURES
source
1..4094
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
368..2000
/gene="mpcA"
368..373
/gene="mpcA"
382..387
/feature="putative"
382..387
/gene="mpcA"
399..404
/feature="putative"
399..404
/gene="mpcA"
422..425
/feature="putative"
422..425
/gene="mpcA"
438..2000
/feature="putative"
438..2000
/gene="mpcA"
438..2000
/feature="mpcA"
/codon_start=1

```





```

1680 TACACTTCTGGTACTCGACGGCTGACCGCAACCGGAGCTCGAG 1729
      ::::: ::::: ||||| ::::: ::::: ||||| ::::: |||||
312 pLysThrLeuLeuPheLeuThrTyraSpGluThrGlyGlyPheTyraSpH 329
      ||||| ||||| ::::: ||||| ::::: ||||| ::::: |||||
1730 CAAGACGGCTGCTGCTGATGTCGACGAGACGACGCTTCCTCGACC 1779
      ::::: ::::: ||||| ::::: ||||| ::::: |||||
329 isValProProLeuAla.....ValArgProAspAsnLeuThrTyR 343
      ||||| ||||| ||| ::::: |||||
1780 ACCTGCCCGCGCGCGCGGCGGCTGAACAGACGACGCGCTGCGT 1829
      ||||| ||||| ||| ::::: |||||
344 ThrGluThr.....AlaLysAsnGlyG 351
      ::::: |||||
1830 GGCAGACCGCGCGCGGCGGCGGCTGCAATGCGACACCAAGCGGATAT 1879
      ::::: |||||
351 nLysTyThrLeuHisPheArgLeuGlyGlyArgMetProThrTyR 368
      ::::: ||||| ::::: ||||| ::::: |||||
1880 CCGTTATCGCAACACGCTACGGCTCGCGCGCGGCTGCGGATGACG 1929
      ::::: ||||| ::::: ||||| ::::: |||||
368 alileSerProTySerLysGlyTyrlleGluGlnTyrglyThrAsp 384
      ||||| ||||| ::::: ||||| ::::: |||||
1930 TGATCTCGCGGTGAGACAGCGGCGGCTGCTCAACTCCCA.GGTGTTCCA 1978
      ::::: |||||
385 ProValThrGlyLysProAlaProTySerAlaThrSerValLeuLysTh 401
      ||| ||| ||| ||| ::::: |||||
1979 CCACACCTCGGTGATCGCTTCTCGACGACGCTCGG..... 2017
      |||||
401 rleuGlyTyThrLeuThrAspIleuAspPheThrProArgValAlaHis 418
      ::::: |||||
2018 .....GGTCATG 2024
      ::::: |||||
418 erProSerPheAspHisLeuIleGlyThrThrLeuArgGlu...AspAla 433
      ::::: |||||
2025 GAGCGCAATATCAGTCCCTGGCGTCTGCGGCTGCGGCGGCTGACCTC 2074
      ::::: |||||
434 ProIleAlaLeuLysThrProHisThrPheSerVal 445
      ::::: |||||
2075 GCGGTCACCTTCGCCAACCCGCAACACGACGCGTT 2110
      ::::: |||||

```

seq\_name: gb\_ba2:PSEPLCNA

```

seq_documentation_block: 2932 bp DNA BCT 26-APR-1993
LOCUS PSEPLCNA
DEFINITION P.aeruginosa non-hemolytic phospholipase C (plcN) gene, complete cds.
ACCESSION M30223
VERSION M30223.1 GI:151497
KEYWORDS phospholipase.
SOURCE P.aeruginosa (strain PAOI) DNA.
ORGANISM Pseudomonas aeruginosa
Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 2932)
Ostroff,R.M., Vasil,A.I. and Vasil,M.L.
Cloning expression and nucleotide sequence of the gene encoding a
non-hemolytic phospholipase C from Pseudomonas aeruginosa
J. Biol. Chem. (1989) In press
JOURNAL
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by M.L.Vasil, 30-NOV-1989.
FEATURES
Location/Qualifiers
source 1..2932
/organism="Pseudomonas aeruginosa"
/db_xref="taxon:287"
759..863
/note="non-hemolytic phospholipase C signal peptide"
759..2837
/note="non-hemolytic phospholipase C precursor"
/codon_start=1
/transl_table=11
/protein_id="AA25969.1"
/db_xref="GI:151498"
/translation="MISKRRSRFIRLAAGTAAATVATSMLPSSIOAALAIAPARRHGN
LKDEHVIVILMOENRSDHFYFTLKGVRFGDRMAIPLPDQQRVHKGKSGEILPYH

```

```

FDTSTTSQRVDGTPHTWPDQAQWNEGRMDKWLPAKTERSLGYKKEQDIAFOFAMAN
AFTICDAYHCSFGSGGTNPRLFLMTGTNDPLQGHGGVTTNDHDSNGPVBQGYTWTY
PERLOAGITWRYVQDMADNFDNDPLIGFRQYRAAAPDSPLIVNGLSTWKLALKRDV
LANSLPQSVITVAPAKYSEHPGSPSPIMGAEYTSVLDALTANPEWSEKALLVMFDE
NDGFFDHVAPPAASLNDKDTLRGTATADTLEWHITKDIRYNOPVGLGARVPKVI
SPWSKGMVNSOVSFDHTSVIRELEORFCVMEPNISPRRAVCDDLTSAFNAPNNNEP
EPLEPDTSDADAIVASQIKLPKPPAPVAMPQEMGIRPARALTYELGVHARTRSGG
DALSLTFANTGKAGAVDFDLDSENPPKRYTVGAKRLHDSFOGDASGDYHLEVHG
PNGFLRFRGNLRDLADGKAPLEVIDYELPLGNLRVQLINRGRHPVKLTVDNRY
ROGERRTVNPVPPQORREVRYSLRSSGWNMYDFSVAQADSFLRRFSGRMDGRSGFSD
PGMGLGLTTF"
864..2834
mat_peptide

```

```

BASE COUNT 505 a 1000 c 957 g 470 t
ORIGIN

```

```

alignment_scores:
  Quality: 273.50      Length: 462
  Ratio: 1.205        Gaps: 20
  Percent Similarity: 49.134  Percent Identity: 25.541

```

```

alignment_block:
US-09-426-072-2_COPY_20_464 x PSEPLCNA ..

```

```

Align seg 1/1 to: PSEPLCNA from: 1 to: 2932

```

```

27 AsnLeuLysAspLysValGluAsnIleValTrpLeuIleLeuGluAsnAr 43
||||| |||||
888 AACCTCAAGGAC...GTCGAGCAGCTGGTGATCTGATCGAGGAGAACCG 934
||||| |||||
43 gSerPheAspAsnIleLeuGly.....GlyValArgArgGlnGly. 56
||||| |||||
935 CTCCTTCGACCATATTTCGGGACCCCTCAAGGGCGCTCGCGGCTTCGCGG 984
||||| |||||
57 .....LeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrlLysAsn 71
||| |||||
985 ACCGATGCGCATCCCGCTGCCGATGCCAGCGGGTCTGGCACCAGAAAG 1034
|||||
72 AlaSerAspProSerSerGlyLysTyR.....CysThrG 83
|||||
1035 GGCACGAGGCGGAGATCTCGCCCTACCACCTTCACACACGACGACCCAG 1084
|||||
83 nAlaLysAspTyraSerSerValPheAsnAspProAspHisSerValThrG 100
|||||
1085 CGCCGAGCGCGTCGACGGC..... 1103
|||||
100 LysAsnAsnLeuGluPheTyrglyThrThrProAsnAsnGlyAlaIle 116
|||||
1104 .....ACCCGCGCACACCTGGCGCGGAC 1124
|||||
117 AlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLe 133
|||
1125 GCC.....CAGCAGGCTTGAACGAGGCGGCGAT 1153
|||||
133 uAsnAspTyraProLysLeuAlaProGluGluAlaThrArgGlnValMetG 150
|||||
1154 GGAC.....AAGTGCTCGCGGCGCAAG...ACCGAGCGTTCCTCTGG 1191
|||||
150 lyTyrrTyrrGluGluGluValProThrLeuValAspLeuValAspGlu 166
|||||
1192 GCTACTACAAGAGCAGGACATCGCTTCAGTTCGCGATGGCCCAACGCC 1241
|||||
167 PheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsnPr 183
|||||
1242 TTCACCATCTGCCAGCGCTATCTACGTCTGTCAGGGCGGCGGCAACCC 1291
|||||
183 oAsnArgLeuCysAlaLeuAlaGlyThr.....AlaAlaGlyHisG 197
|||||
1292 CAACCGGCTGTTCTCTGAGCGCGCACCAACGACCGCTCGGCCAGCAGC 1341
|||||
197 Ly.....LysAsnAspAspPhe.....LeuAsnTyrr 206
||| ||||| |||

```

AE004754.1 GI:9949446  
 .  
 Pseudomonas aeruginosa.  
 Pseudomonas aeruginosa  
 Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 Pseudomonas  
 1 (bases 1 to 10092)  
 Stover,C.K., Pham,X.Q., Erwin,A.L., Miziochuchi,S.D., Warrenner,P.,  
 Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,  
 Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,  
 Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,  
 Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.  
 Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
 opportunistic pathogen  
 Nature 406 (6799), 959-964 (2000)  
 20437337  
 2 (bases 1 to 10092)  
 Stover,C.K., Pham,X.Q., Erwin,A.L., Miziochuchi,S.D., Warrenner,P.,  
 Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,  
 Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,  
 Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
 Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,  
 Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,  
 Hancock,R.E.W., Lory,S. and Olson,M.V.  
 Direct Submission  
 Submitted (16-MAY-2000) Department of Medicine and Genetics,  
 University of Washington Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 Location/Qualifiers  
 1..10092  
 /organism="Pseudomonas aeruginosa"  
 /strain="PA01"  
 /db\_xref="taxon:287"  
 93..821  
 /gene="PA3317"  
 93..821  
 /gene="PA3317"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG06705.1"  
 /db\_xref="GI:9949447"  
 887..1387  
 /gene="PA3318"  
 887..1387  
 /gene="PA3318"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG06706.1"  
 /db\_xref="GI:9949448"  
 /transl\_table="MNRFLPRPDAROCVAGLFSALLSSANAAACRDTISVQGS  
 ADATRTSGGHVTHLYGWTPESSQDKTLFKSRGNYDAARQYGYLDNPVACSG  
 NSKFQVSLXELHMGKIDAYSKQADGGACFTWDTYWAKEISYGVFLDKGKLWLTLL  
 YPVLQ"  
 complement(1445..3523)  
 /gene="plcN"  
 /note="PA3319"  
 complement(1445..3523)  
 /gene="plcN"  
 /codon\_start=1  
 /transl\_table=11  
 /product="non-hemolytic phospholipase C precursor"  
 /protein\_id="AAG06707.1"  
 /db\_xref="GI:9949449"  
 /transl\_table="MISKSRRSFIRLAAGTVGTATVATSMLPSSQAALAIPAHRRHGN  
 LKDVHRTILMQEHSDFHFTGLKVRGCDRMAIPLPDGQRVHKGQCEILPHT  
 FDTSTTSAQRVDQHPHTWDAQAQANECRMDKWLPAKTRISLYGKEQDIAQFAPKAM  
 YDVT"

seq_documentation_block:	DNA	BCT
LOCUS	AE004754	10092 bp
DEFINITION	Pseudomonas aeruginosa PA01, section 315 of 529 of the complete genome.	
ACCESSION	AF004754	AE004091



```

AFTICDAYHCSFOGFTNPNRLFLMTGTNDPLGOHGGPVTNDHDSNGPVEQGYTWTY
PERLQAAGITWRYVQDMADNFDNDPLIGFROYEAPAPDPSPLVINGLSTWKLDAKRDV
LANSLOVSNIVAPAKYSEHCPSPPIWGAEYTSWYLDALTANPEVWSKTAALLVMPDE
NCGFEDHVAIPAPSLKNDGTLTKADATLEWHTKGDIRYRNOQYGLGPRVPMYVI
SPWKGWNSQVDFDHTSVIRFLRFORFVMEPNISPMRRVAGDULTSAFNANPNNEP
FPELPDTSQADAIVASQIKLPKPAAVAAKQEMGIIPARALPELGYHARYSRGG
DALSLTANTGKAGAVQVFDLDSENPPKRYTVGARKRLHDSFGDASGDYHLEVHG
PAGFLRVGRNLRDLAERKAPLPEDRIDYEPFLGNLRVQLINRGRHPVKLTVDKNVY
RQGERRTVNVPPQCORREVRYSLRSSGNWYDFSVSAQAGDSFLRRFSGRMEDGRSGFSD
PCMGLGTLTF"
complement(3754..4209)
/gene="PA3320"
CDS
complement(3754..4209)
/gene="PA3320"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG06708.1"
/db_xref="GI:9949450"
/transl_table="MPNLYLMLVAVSGMYVLQAGSNALLGRLLGHPHLSLSLGV
SLGVYSLGULRVPLUPATASLGGAPWWSWLGFLGAAAYLTVALVAPRIGATFIAC
VWAGOMLVSAICDQFGWAGFPVRRLSWEGLYGLALVAVAGVLLQWRGA"
4298..5212
/gene="PA3321"
4298..5212
/gene="PA3321"
/codon_start=1
/transl_table=1
/product="probable transcriptional regulator"
/protein_id="AAG06709.1"
/db_xref="GI:9949451"
/transl_table="MQQIHDLRLIDMLLVLDALLSERHVSRAAORLAMSOPVASHA
LCRLRELFGDPLLVKRSKMDLTLLELAPKLAQSDQVRGVIGGPTPEPRVSRSH
RLGMSDYGAWVLPALLRQQAALPAGVSGFVIOEPRLETAQVAEGOLCALGVPELL
PEGVARRRLEPERFVCVCRSLDPDGLSLDRLYLAARHVSLODNPAEVDLSALEK
BGLRROVALTLPHFTVAPALLAGTDVLVIAETLRLRLPELALCEPLLAIPRFDY
VQIRGDAEREANGAWLREQIQAARKI"
complement(5220..5918)
/gene="PA3322"
complement(5220..5918)
/gene="PA3322"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG06710.1"
/db_xref="GI:9949452"
/transl_table="MRNRYAYQDRGLPPRRGRGVYFDSHLKRLRLISQLERGY
SSANIKELLEAMEQGRDLHVLGDQAIIGENWLETGSGIFGDELQAIIFGDELNDRVI
DKAQLGLLFDGERMKIPSPRFQAGLELYRSGIPLEALLDQLGELRSDTEHLAAGI
VRLIVTHLVDRGTDLLFGAPDLQELAKVFLRLPRLAEHVQVVELARGLKLKANEMIG
ERVCEMLRLEKPS"
6344..7219
/gene="PA3323"
6344..7219
/gene="PA3323"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG06711.1"
/db_xref="GI:9949453"
/transl_table="MSQANVYPPDDLDHFRQVQRLATTCETIHGELRPGISEREA
ALMTWLQDHGVDDFHQFPFAMGSRISFSGKDFGLRHGLGFNLAIFYGKORLEEGMPF
ILDCVITGAYTADYGVSGALGNHNLQERLMDLIIHRRITLIDLVRERPLAEVSOAV
DRLCVTOGTLPRKHAYPEVLAHRVEKLGKRNRPSPVARGVNRNLYTLTNALVTKRE
GWSPLWSGKRSEHPPTSGLWAVEPHLGLRGVGAKEEELVVTEDDAYWLDLDLPHVR
RWQARGLLPQAREAA"
7216..8994
/gene="PA3324"
7216..8994
/gene="PA3324"
/codon_start=1
/transl_table=1
/product="probable short-chain dehydrogenase"

```

```

/protein_id="AAG06712.1"
/db_xref="GI:9949454"
/transl_table="MNAALIALPDQLPERLEVQSGDVRVLAVHWCAPONGKPTLLMVHG
YPDHETWPLIIRQLAGRYRIVAYDVRGAGASDKPRMSRDYHLQRLSEDLOAVIRATS
ALRPVQLLAHDWGSIOFTNESVTDPQCPPLIASYTSISGCLDHDVGVWNRHLPORSPK
PKRVAFFQQLLHSHYIAFFHTVPVPELLASLAGLWLPOLKHAEGVRHPOVNPQASD
GRHGKLYRCNFIRSLFPRKRHTEVPOLIVPTRORYYQALQFQHLSLWAPRLMRE
ASVGHMOLLAEPQLAGLWGEFIDAQETGSPPLAQRAQVRPDASSMNGKLVVVYTCAG
GGIRSLTSLSFACRGLASLLAADLSAEERSAELARAGATAHVYQVYDQDTQAMETF
AEMVRFLGVPDVVVVSNAGIMAGPMIDTSPAEMERLLRVNLWSVIDGCRLFQROMIA
ANKPGLHVNVSAGVAFAPSRNPAYATSKAAVLMSECLRAELAGRSIGVTAVCPGFV
DTGIVQATREVGMDAERQARQAKIQIRFYRRRLSPDTVGEKLVRAVERNKAUVSVGS
EYHLGALQWRFPAPWATREFLAREDLTS"
9004..9891
/gene="PA3325"
9004..9891
/gene="PA3325"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG06713.1"
/db_xref="GI:9949455"
/transl_table="KSEPHKLQVKYRDFDSTPLHWVPGSEASOVMMVLMHLLPS
GELNFCRLYNKALPLVSDARLDDVQGFIRQEAHMSRAHDGAIQIRYLSRFDIDSKPFT
DKMDFLFRLLCDYPLUGNRLTRALDGMWRQRCATIAAVEHETGVLGKWLEARELD
RRGSDPYMLDLLRHWGAEVEHRSVAHDLHVLGGSFVTRQLTMLLVIALVLYLEADG
TRTLDDQDPSLRGAKRQGLRTMWRMLGKQGLLPNLARTFROVARYFNPNYPHPEANT
OEALLEYLANSPPAARAANA"
BASE COUNT 1686 a 3490 c 3309 g 1607 t
ORIGIN

alignment_scores:
  Quality: 271.50 Length: 462
  Ratio: 1.201 Gaps: 20
  Percent Similarity: 48.918 Percent Identity: 25.541

alignment_block:
US-09-426-072-2_Copy_20_464 x AE004754/rev ..
Align seg 1/1 to reverse of: AE004754 from: 1 to: 10092

27 AsnLeuTaspLysValGluAsnLleValTrpLeuIleLeuGluAsnAr 43
|||||
3394 AACCTCAAGGAC...CTCAGCAGCGTGTGATCTGATCGAGGAGAACCG 3348

43 gSerPheAspAsnIleLeuGly.....GlyValArgArgGlnGly. 56
|||||
3347 CTCCTTCGACCACCTATTTCGGGACCCCTCAAGGCGTCCCGCGCTTCGGCG 3298

57 .....LeuAspAsnProIleAsnAsnGlyProPheCysAsnTyrIleAsn 71
|||||
3297 ACCCGATGGCGGATCCCGCTGCCGCGGCGCAGCGGCTCTGCGCAGCAAG 3248

72 AlaSerAspProSerSerGlyLysTyr.....CysTrpG1 83
|||||
3247 GCGACCAAGCGCGAGATCTGCCCTACCCTTCACACACAGCAGCAGCAG 3198

83 nAlaLysTyrPaspSerValPheAsnAspProAspHisSerValThrG 100
|||||
3197 CGCCACGCGCTCGCAGGG..... 3179

100 lYAsnAsnLeuGluPheTyrGlyThrTyrThrProAsnAsnGlyAlaIle 116
|||||
3178 .....ACCCCGCACACCTGGCCCGAC 3158

117 AlaSerGlyLysValValAlaAspGlnSerGlyPheLeuAsnAlaGlnLe 133
|||||
3157 GCC.....CAGCAGGCGCTGCAACGAGGCGCAT 3129

133 uAsnAspTyrProLysLeuAlaProGluGluAlaThrArgGlnValMetG 150
|||||
3128 GGAC.....AGTGGCTGCCGCCAAG...ACCGAGGCTTCCCTGG 3091

```

```
150 lyTyrThrGluGluValProThrLeuValAspLeuValAspGlu 166
||||| |||:||||:
3090 GCTACTACAAGGACGAGCATCGCTTCCAGTTCGGGATGCGCAACGCC 3041
||||| |||:||||:
167 PheThrThrPheAsnSerTrpPheSerCysValProGlyProThrAsnPr 183
||||| |||:||||:
3040 TTCACCATCTCGGACGGCTATCACTGCTGTTCCAGGGGGGACCAACCC 2991
||||| |||:||||:
183 oAsnArgLeuCysAlaLeuAlaGlyThr.....AlaAlaGlyHisG 197
||||| |||:||||:
2990 CAACGGCTGTTCTCTGGACCGGACCAACGACCGCTCGGCCAGCAG 2941
||||| |||:||||:
197 ly.....LysAsnAspAspPhe.....LeuAsnTyr 206
|| ||||| |||
2940 GTGGTCCGCTAACCAACGACGACGACGACGACGCGCGGTGGAGGAG 2891
||||| |||:||||:
207 GlyIleSerSerLysSerIlePheGluAlaAlaAsnGluLysGlyVal 223
||| |||: |||
2890 GGCTACACCTGGACCACTATCCGAGCGCTGCAGGCTCGCGGCATCAC 2841
||| |||: |||
223 rTrpLeuAsnTyrAspGlyThrAsnGlyClupheGluProAspSerLeu. 239
||||| |||: |||
2840 CTGGGGGCTTACGAGACATGCGGACAACTTCTCCGACACCCGCTGA 2791
||||| |||: |||
240 ....PhePheThrTyrValAsnGlnThrSerArgSerAsnVal... 253
||| ||| |||
2790 TCGGCTTCGGCAGTACCGCGCGCGCGCTGACTCGCGCTGATCGTC 2741
||||| |||: |||
254 .....ProValGluAsnPheGlnAspAlaTyrLe 264
||||| |||: |||
2740 AACGGCTGAGCACTCGAAGCTCGATCGCTGAAGCGGAGCTGCTGCG 2691
||||| |||: |||
264 uGlyValLeuProLysPheSerTyrIleAsnProSerCysCysGlyThr 281
||| ||||| |||
2690 CAATAGCTCGCGCAGGTGCTCTGGATC.....GTGCGCCCGG 2653
||||| |||: |||
281 snThrAsnSerMetHisPro...ThrGlyAsnValSerTyrGlyGluVal 296
||| ||||| |||
2652 CCAAGTACTCGAAGACCCCGCGCGCTCCAGCCGCTGTTGGGGGCGGAG 2603
||||| |||: |||
297 PheValIleGlnIleTyrAspAlaIleArgGlnGlyProGln...TrpAs 312
||||| |||: |||
2602 TACACTTCTGGTACTCGACGGCTGACCGCAACCGGAGTCTGGAG 2553
||||| |||: |||
312 pLysThrLeuLeuPheIleThrTyrAspGluThrGlyGlyPheTyrAspH 329
||||| |||: |||
2552 CAGACCGCGCTGCTGGTGATGTCGACGAGACGACGCGCTCTTCGACC 2503
||||| |||: |||
329 isValProProLeuAla.....ValArgProAspAsnLeuThrTyr 343
||||| |||: |||
2502 ACCTCGCGCGCGCGCGCGCTGAACAAGACGCGCGCTGCGT 2453
||||| |||: |||
344 ThrGluThr.....AlaLysAsnGlyI 351
||||| |||: |||
2452 GGAAGACACCGCGCGCGCGCGCTCGAATGGCACACGAGGGCGATAT 2403
||||| |||: |||
351 nLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTrpV 368
||||| |||: |||
2402 CCGTTATCAACCAACCGCTACGGCTCGCGCGCGCGGTCCGATGTAGC 2353
||||| |||: |||
368 alIleSerProTyrSerLysGlyTyrIleGluGlnTyrGlyThrAsp 384
||||| |||: |||
2352 TGATCTCGCGTGGGCAAGCGCGCTGGGTCAACTCCCA.GGTGTTCGA 2304
||||| |||: |||
385 ProValThrGlyLysProAlaProTyrSerAlaThrSerValLeuLysTh 401
||||| |||: |||
2303 CCACACTCGGTGATCGCTTCCTGGAGCAGCGCTCGG..... 2265
||||| |||: |||
401 rLeuGlyTyrLeuThrAspIleGluAspPheThrProArgValAlaHis 418
||||| |||: |||
2264 .....GGTCATG 2258
```

```
418 erProSerPheAspHisLeuIleGlyThrThrLeuArgGlu...AspAla 433
||| |||: |||
2257 GAGCGGAATATCAGTCCCTGGCGTCTGCGGCGACCTGACCTC 2208
||||| |||: |||
434 ProIleAlaLeuLysThrProHisThrPheSerVal 445
||| |||: |||
2207 GCGTTCAACTTCGCCAACCCGACACACGAGCGGTT 2172
```

OM of: US-09-426-072-2\_COPY\_20\_464 to: EST.\* out\_format : pfs  
 Date: Feb 27, 2001 2:34 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=frameat\_p2n\_model -DEV=slp  
 -O=/cgn2\_1/USPTO.spool/6146869/runat\_13022001\_095136\_14969/app\_query.fasta\_1.1049  
 -DB=EST -QPMF=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
 -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
 -ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -MINLEN=0  
 -MAXLEN=200000000 -USER=6146869 -CENL\_1\_3184 -NCPU=6 -ICPU=3  
 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-426-072-2\_COPY\_20\_464  
 Query length: 445  
 Database: EST.\*  
 Database sequences: 7991742  
 Database length: 79122348  
 Search time (sec): 1306.080000

## score\_list:

Sequence	Strd	Orig	EScore	Len	Documentation
gb_est46:AW565749	+	439.50	887.03	2.1e-40	1. Documentations
gb_est75:BE758821	+	311.50	623.43	1.0e-25	1. AW565749 LG1_349_G03.gi_A002
gb_est66:BE053545	+	303.00	600.52	1.9e-24	1. BE758821 an_3066 Aspergillus ni
gb_est48:AW726682	+	286.00	567.84	1.3e-22	1. BE053545 GA_Ea0035B08f Gossyp
gb_est70:BE404795	+	282.50	562.07	2.7e-22	1. AW726682 GA_Ea0022G10 Gossyp
gb_est47:AW671667	+	278.50	556.23	7.4e-22	1. BE404795 WHE1205_E03_1052s Whea
gb_est51:AW929343	+	254.50	504.37	4.4e-19	1. AW671667 LG1_349_G03.gi_A002
gb_est78:BE000464	+	253.50	504.09	4.6e-19	1. AW929343 EST338131 tomato flowe
gb_est23:BA1691435	+	253.50	501.60	6.3e-19	1. BF004644 EST433142 KVI Medicago
gb_est71:BB436321	+	244.50	484.70	5.5e-18	1. A1691435 606015H11 xl 606 - Eay
gb_est78:BF0004645	+	242.50	480.39	9.5e-18	1. BB436321 EST407399 tomato break
gb_est78:BF010466	+	241.00	481.44	8.4e-18	1. BF010466 NCX1_084_COL_F NCX1 (N
gb_est59:AA556789	+	236.50	465.75	6.2e-17	1. AA556789 631 Lobiolly pine NA F
gb_est24:AW266012	+	234.50	463.16	8.7e-17	1. AW266012 L30-2674T3 Ice plant I
gb_est48:AW750470	+	229.50	454.85	2.5e-16	1. AW750470 s121g05.y1 Gm-c1036 G
gb_est27:AI994923	+	228.00	451.81	3.7e-16	1. AI994923 701500427 A. thaliana
gb_est73:BE586656	+	227.50	451.15	4.1e-16	1. BE586656 WHE0509_A09_A172R Seca
gb_est41:AW171686	+	226.50	446.83	7.1e-16	1. AW171686 N100580a rootphos(-) M
gb_est42:AW266415	+	226.00	447.84	6.2e-16	1. AW266415 L30-3180T3 Ice plant I
gb_est66:BE057023	+	222.50	440.25	1.6e-15	1. BE057023 945037H04.y1 945 - Mix
gb_est41:AW224301	+	221.50	439.23	1.9e-15	1. AW224301 EST301028 tomato fruit
gb_est70:BE415670	+	216.50	428.48	7.4e-15	1. BE415670 MML037_H05000424 ITEC
gb_est85:TA4962	+	199.00	394.63	5.7e-13	1. TA4962 8225 Lambda-PRL2 Arabid
gb_est29:AU108508	+	185.00	358.99	5.5e-11	1. AU108508 AU108508 Avicennia mar
gb_est79:CB26175	+	181.50	357.68	6.5e-11	1. CB26175 C26175 Rice callus CDNA
gb_est48:AW730355	+	180.00	344.27	3.6e-10	1. AW730355 GA_Ba0023E24 Gossyp
gb_est74:BE662652	+	178.00	330.91	1.6e-10	1. BE662652 EST90590F09 Pine triph
gb_est70:BE343260	+	177.50	344.61	3.5e-10	1. BE343260 EST396092 potato stold
gb_est29:AU065663	+	172.00	335.37	1.1e-09	1. AU065663 AU065663 Rice cDNA fro
gb_est70:BE354915	+	171.50	332.16	1.7e-09	1. BE354915 EST355005 tomato flowe
gb_est26:BA1896182	+	163.50	317.17	1.2e-08	1. BA1896182 EST256525 tomato callu
gb_est71:BB430717	+	161.00	306.70	4.5e-08	1. BB430717 SUN007_E02F991221 ITEC
gb_est68:BE187301	+	160.50	308.94	3.4e-08	1. BE187301 NKNW_162_C04_F Nsf Xyl
gb_est29:AU032411	+	155.50	301.98	8.3e-08	1. AU032411 AU032411 Rice root Ory
gb_est17:AI162187	+	152.00	291.33	3.2e-07	1. AI162187 AU13P400 Hybrid aspen
gb_est36:AU093977	+	146.50	284.43	7.9e-07	1. AU093977 AV409377 Lotus Japonic
gb_gss14:AO969055	+	145.00	278.22	1.7e-06	1. AO969055 LERJ170TR LERG Arabid
gb_est24:AI776678	+	142.00	271.02	4.4e-06	1. AI776678 EST257790 tomato resis
gb_est39:AW043326	+	139.00	264.61	1.0e-05	1. AW043326 ST32A07 Pine Triple s
gb_gss14:AO969054	+	138.00	264.57	1.0e-05	1. AO969054 LERJ170TF LERG Arabid

gb\_est42:AW289822 + 136.50 263.85 1.1e-05 414 | AW289822 NXNV006G10F Nsf Xyl  
 gb\_est28:AL369844 + 131.00 250.63 6.0e-05 480 | AL369844 MLBA33G10F1 MlBA Me  
 gb\_est82:H76814 + 130.00 248.26 8.1e-05 492 | H76814 17245 Lambda-PRL2 Ara  
 gb\_est27:AI992399 + 122.00 229.75 0.0009 575 | AI992399 701557637 A. thalia

seq\_name: gb\_est46:AW565749

seq\_documentation\_block: 667 bp mRNA 19-JUL-2000  
 LOCUS AW565749  
 DEFINITION LG1\_349\_G03.gi\_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA  
 sequence.  
 ACCESSION AW565749  
 VERSION AW565749  
 KEYWORDS EST.  
 SOURCE EST.  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poales; Poaceae; Sorghum.  
 REFERENCE 1 (bases 1 to 667)  
 AUTHORS Cordonnier-Pratt,M.,M., Gingle,A., Marsala,C. and Pratt,L.H.  
 TITLE An EST database from Sorghum: light-grown seedlings  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: T7  
 High quality sequence start: 73  
 High quality sequence stop: 667  
 POLYA=Yes.

FEATURES  
 Location/Qualifiers  
 1..667  
 source  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Light Grown 1 (LG1)"  
 /note="Organ: 10- to 14-day-old light-grown (greenhouse)  
 seedlings; Vector: Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI  
 ; The library was made from poly-A RNA in the cloning  
 vector lambda ZAP II. Clones to be sequenced were  
 prepared by mass excision."  
 BASE COUNT 140 a 224 c 173 g 130 t  
 ORIGIN

alignment\_scores:  
 Quality: 439.50 Length: 192  
 Ratio: 3.117 Gaps: 3  
 Percent Similarity: 73.438 Percent Identity: 44.792

alignment\_block:  
 US-09-426-072-2\_COPY\_20\_464 x AW565749

Align seg 1/1 to: AW565749 from: 1 to: 667

250 SerAsnValProValGluAsnPhePheGlnAspAlaTyrLeuGlyVa 266

12 ACCACAGTTCGCCCTGCAGGATTCVACCGGACGCCCAACGGCAA 61

266 lleuPolysPheSerTyrIleAsnProSerCysGlyThrAsnThrA 283

62 TCTCCCGCAGTTCACGTGATCAACCCGAGTGTGCG.....TCGTACC 105

283 snSerMetHisProThrGlyAsnValSerTyrGlyGluValPheVallys 299

106 AGTCCTCCACCCCGCGTCTCCCATTCGACGAAGCGAAAGTTCTCAAG 155

300 GlnIleTyrAspAlaIleargGlnGlyProGlnTrpAspIysThrLeuLe 316

RNA isolated from *Aspergillus niger* grown in both complete

of the cotton fiber

of the cotton fiber

REFERENCE	AUTHORS	TITLE
1	Leslie, A., Frisch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A.	An integrated analysis of the genetics, development, and evolution of the cotton fiber

JOURNAL  
COMMENT

unpublished (2000).  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Tel: 864 656 4293  
Fax: 864 656 4298  
Email: [rwing@clemson.edu](mailto:rwing@clemson.edu)  
Seq primer: TAATGACCTACTATAGG  
High quality sequence start: 2  
High quality sequence stop: 78

FEATURES  
source

FEATURES	source
high quality sequence stop: 720.	
Location/Qualifiers	
1. 865	
/organism="Gossypium arboreum"	
/strain="AKA"	
/cultivar="8400"	
/db_xref="taxon:29729"	
/clone_lib="GA_Ea0035B08f"	
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"	
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"	
/lab_host="E. coli"	
/notes="Vector: pBK-CMV; Site_1: ECORI; Site_2: XhoI"	
242 a	205 c 191 g 224 t 3 others
BASE COUNT	

alignment\_scores:

Quality:	303.00	Length:	290
Ratio:	1.859	Gaps:	12
Percent Similarity:	56.207	Percent Identity:	32.414

**alignment\_block:**

US-09-426-072-2\_COPY\_20\_464 x BE053545

Align seq 1/1 to: BE053545 from: 1 to: 865

[illegible]

```

317 PheileThrTyraSpGluThrGlyGlyPheTyraSpHisValProProPr 333
      : : : : : : : : : : : : : : : : : : : : : : : : : :
433 CTGGTCATATGATCAACATGGTGGTTCTATGACCATGTTCCGACACC 482
      : : : : : : : : : : : : : : : : : : : : : : : : : :
333 oLeuAlaVal...ArgProAspAsnLeuThrTyThrGluThrAlaLysA 349
      | : : : : : : : : : : : : : : : : : : : : : : : :
483 ACCGGAGTCCCTAGCCCGATGATATGTCGGTCTCTGAG..... 522
      : : : : : : : : : : : : : : : : : : : : : : : : : :
349 snGlyGlnLysTyThrLeuHisPheAspArgLeuGlyGlyArgMetPro 365
      | : : : : : : : : : : : : : : : : : : : : : : : :
523 .....CCTTATACTTCAAGTTTGATCGTCTGTTGCAGGGTTCTCT 564
      : : : : : : : : : : : : : : : : : : : : : : : : : :
366 ThrTrpValIleSerProTyraSer..... 373
      : : : : : : : : : : : : : : : : : : : : : : : : : :
565 GCATATGGTTTCCCTTCCTGATGAGCTGGACACGTGTCATAGGC 614
      : : : : : : : : : : : : : : : : : : : : : : : : : :
374 .....LysLysGlyTyrIleGluGlnTyrGlyThrAspProValThrG 388
      : : : : : : : : : : : : : : : : : : : : : : : : : :
615 CATCAGGGCCAGATTCTACATC.....A 637
      : : : : : : : : : : : : : : : : : : : : : : : : : :
388 lYlysProAlaProTySerAlaThrSerValLeuLysThrLeuGlyTy 404
      ||| : : : : : : : : : : : : : : : : : : : : : : : :
638 NAGNTCGAGCATCTCCNCAATTGCAGCAACACTTATAG..... 675
      : : : : : : : : : : : : : : : : : : : : : : : : : :
405 LeuTrpAspIleGluAspPhe...ThrProArgValAlaHisSerProSe 420
      : : : : : : : : : : : : : : : : : : : : : : : : : :
676 ATTTTCAATCTCAACAGATTTTCTAACCAACGCGTGATCGCTGGCGTGG 725
      : : : : : : : : : : : : : : : : : : : : : : : : : :
420 rPheAspHisLeuIleGly.....ThrThrLeuArgGluAspAlaP 434
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :
726 CTTTGTAT.....ATTGGTGGCAATCGAGGCCACCCCAAGACACTGTG 769
      : : : : : : : : : : : : : : : : : : : : : : : : : :
434 roileAlaLeuLysThrPro 440
      ||| : : : : : : : : : : : : : : : : : : : : : : : :
770 CCGAAAAACTGGCAAAACCC 789
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq\_name: gb\_est48:AW726682

seq\_documentation\_block: 698 bp mRNA EST 20-APR-2000  
LOCUS AW726682 Gossypium arboreum 7-10 dpa fiber library Gossypium  
DEFINITION GA\_EA022G10 Gossypium arboreum cDNA clone GA\_EA022G10, mRNA sequence.  
ACCESSION AW726682  
VERSION AW726682.1 GI:7624221  
KEYWORDS EST.  
SOURCE Gossypium arboreum.  
ORGANISM Gossypium arboreum.

## REFERENCE

REFERENCE 1 (pages 1 to 958)  
AUTHORS Leslie, A., Frisch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A.  
TITLE An integrated analysis of the genetics, development, and evolution  
of the cotton fiber  
JOURNAL Unpublished (2000)  
COMMENT Unpublished (2000)  
CONTACT Contact: Wing RA

FEATURES  
SOURCE

```

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"

```

/lab\_host="E. coli"  
 /note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 196 a 155 c 156 g 187 t 4 others  
 ORIGIN

alignment\_scores:  
 Quality: 286.00 Length: 220  
 Ratio: 2.119 Gaps: 9  
 Percent Similarity: 61.364 Percent Identity: 34.091

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x AW726682

Align seg 1/1 to: AW726682 from: 1 to: 698

```

236 ProAspSerLeuPhePheThrTyrValAsnGlnThrSerArgSerAsnVa 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 CCATCTACGCTTTTACAGGAAGCTTAGG.....CACTT 63
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 lValProValGluAsnPhePheGln.....AspA 262
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 GAATATGTGGACATTTCCATCAATACGATCAAGCTTCACGCGTCACT 113
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
262 laTyrLeuGlyValLeuProLysPheSerTyrIleAsnProSerCysCys 278
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 GTAAGGATGGGAAGCTACCAAAATTATGTGTTGATTGAGCCAGATATTTT 163
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 GlyThrAsnThr.....AsnSerMethHisProThrGlyAsnValSe 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 GACATTTTAAACAGCTGTGCAACAGCAGCACCCTTCCCATGACGTCTC 213
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 rTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgGlnGlyP 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 AGAGGCCAGCAAGCTTGTGAAGAAATCTATGAAGCGCTCAGATCAAGTC 263
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 roGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyGly 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 CTGAGTGGATGAATCTTGTTCCTGGTCATATATGATGATGATGATGATG 313
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326 PheTyrAspHisValProProProLeuAlaVal...ArgProAspAsnLe 341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 TTCTATGACCATGTTCGCGACACCAACCGAGTCCCTAGCCCGCATGATAT 363
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 uThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrThrLeuHisPheA 358
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 TGTCGCTCTCTGAG.....CCTATAACTTCAAGTTTG 395
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
358 spArgLeuGlyClyArgMetProThrTrpValIleSerProTyrSerLys 374
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
396 ATCGTCTTGGTTGAGGGTTCCTGCGCATATGTTTCCCTTGGATTGAG 445
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 LysGlyTyrIle.....GluGlnTyrGlyThrAspProValThrGlyLy 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
446 CCTGGAACAGTGTTCATAGGCATCAGGGCCAGATCCTACA..... 487
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 sProAlaProTyrSerAlaThrSerValLeuLysThrLeuGlyTyrLeuT 406
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
488 ....TCAGATTCAGACATTCCTCCATTGCAGCAACACATTAAGACAGATT 533
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 rpAspIleGluAspPhe...ThrProArgValAlaHisSerProSerPhe 421
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
534 TCAATCTCAAGAATTTCTTAACAAAGCGTGATGCGTGGGCTGTTCCCTTT 583
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
422 AspHisLeuIleGly...ThrThrLeuArgGluAspAlaProIleAlaLe 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
584 GATATTGTTCATCGAAGCACCACCAAGACAGACTGTCCCAAGAAACT 633
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
437 uLysThrPro 440
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
634 GCGAGAGCCA 643
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq\_name: gb\_est70:BE404795

seq\_documentation\_block:

LOCUS BE404795 617 bp mRNA EST 21-JUL-2000  
 DEFINITION WHE1205\_E03\_1052S wheat etiolated seedling root cDNA library  
 Triticum aestivum cDNA clone WHE1205\_E03\_105, mRNA sequence.  
 ACCESSION BE404795  
 VERSION BE404795.1 GI:9364263  
 SOURCE EST.  
 ORGANISM bread wheat.  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.  
 REFERENCE 1 (bases 1 to 617)  
 AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han  
 ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,  
 Seaton,C.L. and Tong,J.C.  
 TITLE The structure and function of the expressed portion of the wheat  
 genomes  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Olin Anderson  
 US Department of Agriculture, Agriculture Research Service, Pacific  
 West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 51053595773  
 Fax: 51053595818  
 Email: oanderson@nps.usda.gov  
 Sequence have been trimmed to remove vector sequence and low  
 quality sequence with phred score less than 20  
 Seq primer: Strategene SK primer.

FEATURES  
 source

```

1..617
  /organism="Triticum aestivum"
  /cultivar="Chinese Spring"
  /db_xref="taxon:4565"
  /clone="WHE1205_E03_105"
  /clone_lib="Wheat etiolated seedling root cDNA library"
  /tissue_type="Root"
  /dev_stage="Five day old etiolated seedling"
  /lab_host="E. coli SOLR"
  /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
  Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
  , germinated and grown aseptically in the dark at room
  temperature on filter paper with water, nystatin and
  cefotaxime in covered crystallization dishes. Roots were
  harvested. The tissue, total RNA, and poly(A) RNA were
  prepared, a cDNA library was made, and the cDNA clones
  were in vivo excised to give plasmid clones in the
  T3 Clone lab (Choi, Close, Fenton) at the University of
  California, Riverside. Plasmid DNA preparations and DNA
  sequencing were performed in the OD Anderson lab (all
  other authors)."
```

BASE COUNT 141 a 202 c 151 g 123 t  
 ORIGIN

alignment\_scores:  
 Quality: 282.50 Length: 216  
 Ratio: 2.108 Gaps: 8  
 Percent Similarity: 62.037 Percent Identity: 33.796

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x BE404795

Align seg 1/1 to: BE404795 from: 1 to: 617

```

202 AspPheLeuAsnTyrGlyIleSerSerLysSerIlePheGluAlaAs 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 GACCTCATCAAC...GGCTTCCCGCAGAGACCATCTTCGACGCTCGA 68
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 nGluLysGlyValSerTrpLeuAsnTyrAspGlyThrAsnGlyGluPheG 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 GGAGGACGGCTCTCTCTC.....GGCATCTACTACGAGACA 106

```

```

235 luProAspSerLeuPheThrValAlaGlnThrSerArgSerAsn 251
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 TCCCGCGCCAGCTCTTCTACAGAGCCCTCCGAGGCTCAAGCAC..... 150
252 ValValProValGluAsnPhePheGln.....As 261
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 .....CTCTCAAGTTCCACAGCTACAGCCTCAGGTCCTCAAGCTCGA 191
261 pAlaTyrLeuGlyValLeuProLysPheSerTyrIleAsnPro..... 275
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 CCGCGCCAGGGCAAGCTCCCAATTACCTGCTCATCGAGCAGAGCTACT 241
276 ..SerCysGlyThrAsnThrAsnSerMetHisProThrGlyAsnVal 291
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 TCGATTTCAGAGGATTTCGCGCCACAGCAGCAGCAGCAGCAGCTC 291
292 SerTyrGlyGluValPheValLysGlnIleTyrAspAlaIleArgGlnG 308
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 GCCAGGGCCAGAGGTTTGTCAAGAGGCTCTACGAGACACTCGCGGCGAG 341
308 yProGlnTrpAspLysThrLeuLeuPheIleThrTyrAspGluThrGlyG 325
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 CCGCGAGTGAAGCAGCAGCGCTCATCATCACTATGATGAGCATGGTG 391
325 lyeTyrAspHisValProProLeuAlaVal...ArgProAspAsn 340
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392 GCTTCTATGACCATGTTCCAGCGCTGTCAAGTGCCTCAGCCTGATCGG 441
341 LeuThrTyrThrGluThrAlaLysGlnGlnLysTyrThrLeuHisPhe 357
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
442 ATCAATT.....GGCCCTGAGCCCTTACTTCAAGTT 473
357 eAspArgLeuGlyGlyArgMetProThrTrpValIleSerProTyrSerL 374
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
474 TGATCGCTCGGGGTGGCGTCCAGCTTCTCATCTCGCCCTGGGTGG 523
374 yLysGlyTyrIleGluGlnTyrGlyThrAspProValThrGlyLysPro 390
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 AGAAGGCGCCGTCATCATGAACCAAAATGGTCT.....CAGGAAC 567
391 AlaProTyrSerAlaThrSerValLeuLysThrLeuGlyTyrLeuTrp 406
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
568 TCACAATATGAGCATTCATCCATCCCTCGCAGCATTAAGAGCTATT 615

```

seq\_name: gb\_est47.AW671667

```

seq_documentation_block:
LOCUS      AW671667          504 bp      mRNA      EST      19-JUL-2000
DEFINITION LG1_349.G03.b1.A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION  AW671667
VERSION    AW671667.1 GI:7535561
KEYWORDS   EST.
SOURCE     sorghum.
ORGANISM   Sorghum bicolor
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
REFERENCE  1 (bases 1 to 504)
            Cordonier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
            An EST database from Sorghum: light-grown seedlings
            Unpublished (2000)
            Contact: Cordonier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 500

```

```

FEATURES             Location/Qualifiers
     source            1..504
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LG1)"
                     seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                     ; The library was made from poly-A RNA in the cloning
                     vector lambda Zap II. Clones to be sequenced were
                     prepared by mass excision.
BASE COUNT          101 a 175 c 139 g 89 t
ORIGIN
alignment_scores:
    Quality: 278.50      Length: 168
    Ratio: 2.401         Gaps: 7
    Percent Similarity: 69.048      Percent Identity: 37.500
alignment_block:
US-09-426-072-2_COPY_20_464 x AW671667 ..
Align seg 1/1 to: AW671667 from: 1 to: 504

```

```

49  LeuGlyGly...ValArgArgGlnGlyLeuAspAsnProIleAsnAsnG 64
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13  CTTGGGGGTCTCGTCTACCGCAACACATCAGGGCTCGTGGTCGACGC... 60
64  yProPheCysAsnTyrLysAsnAlaSerAspProSerSerGlyLysTyrC 81
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61  .GAGTTCTGCACCCCTGTCAACATCAGGAGCCCTCCAGGGTCAGATCT 109
81  yThr.....GlnAlaLysAspTyrAspSerValPheAsnAspProAsp 95
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 GCGCGCTCAGCAGGCAACCAACGTCGACCT.....CAGGATCCCAAC 153
96  HisSerValThrGlyAsnAsnLeuGluPheTyrGlyThrTyrThrProAs 112
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 CACAGCGTCAGCGCGCGCAACCTGCGAGCTGTTCCAGCACCTTCCACCCGA 203
112 nAsnGlyAlaIleAlaSer.....GlyLysValValAla 124
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 CAGGCGCACCGTCCGCGCACCGCATGCCCTCGCGCGG...TACCAGA 250
124 spGlnSerGlyPheLeuAsnAlaGlnLeuAsnAspTyrProLysLeuAla 140
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 ACATGGCGGCTTCGTCGAGGAGCAGCTCGCGTGTGTAC...AAGACGACA 297
141 ProGluGluAlaThrArgGlnValMetGlyTyrTyrThrGluGluGluVa 157
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 GACTCGGCCAAGTCAGCAGGTGATTGACTACTATCCGCGCGCAGCAGCAT 347
157 lProThrLeuValAspLeuValAspGluPheThrThrPheAsnSerTrp 174
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 CCCCCTTCTCGCGCTGGTCGAGAACTTTTGTCTCTTTGACCACCTGT 397
174 heSerCysValProGlyProThrAsnProAsnArgLeuGluCysAlaLeuAla 190
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
398 TCGCCTCTGTGCGAGGCGCGCAACCAACCGGACCGTCCGTACTCGACGTCC 447
191 GlyThrAlaAlaGlyHisGlyLysAsnAspAspPheLeuAsnTyrG 207
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 GGCACAGTCCCGTCTATGGCAGCAACGATAACGCTTTCAACGCTGACGG 497
207 yIle 208
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
498 ACTC 501

```

seq\_name: gb\_est51.AW929343

```

seq_documentation_block:
LOCUS      AW929343          596 bp      mRNA      EST      30-MAY-2000

```

DEFINITION EST338131 tomato flower buds 8 mm to pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone cT0C7I18 5', mRNA sequence.

ACCESSION AW929343

VERSION AW929343.1 GI:8104744

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 596)

AUTHORS van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.

TITLE Generation of ESTs from tomato flower tissue

JOURNAL Unpublished (1999)

COMMENT Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@clemson.edu

5 prime sequence.

# FEATURES

source

1..596

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone\_lib="cT0C7I18"

/tissue="flower"

/dev\_stage="buds 8mm to pre-anthesis, Cornell University"

/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA96). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

BASE COUNT 172 a 127 c 129 g 168 t

ORIGIN

alignment\_scores:

Quality: 254.50 Length: 188

Ratio: 2.194 Gaps: 7

Percent Similarity: 61.702 Percent Identity: 33.511

## alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x AW929343 ..

Align seg 1/1 to: AW929343 from: 1 to: 596

262 AlATyrlGluValLeuProLysPheSerTyr..... 272

2 GCTAAACTTGGCAACATTCACGTTGTGATGTAACAGAGATATTT 51

273 .....ileasProSerCysGlyThrAsnThrAsnSerMetH 286

52 CGATGTTAAATTAACCCAGCA.....ATGATGATC 83

286 isProThrGlyAsnValSerTyrGlyGluValPheValLysGlnIleTyr 302

84 ATCCATCACACAGTAGTACACTTGGACAAAATTTGTTAAGAGGTTTAT 133

303 AsPAIaileArgGlnGlyProGlnTrpAspLysThrLeuLeuPheIleTh 319

134 GAGACTTTCAGGCAAGTCCACAGTGAAGAAATGCGACTCTTGATTAC 183

319 rTyAspGluThrGlyGlyPheTyAspHisValProProLeuAla. 335

184 ATATGACGAACATGTGGCTCTTTGATCATGTACCAACCTCCTGTTCTG 233

336 .....ValArgProAspAsnLeuThrTyrThrGluThrAlaLysAsnGly 350

234 GTGTGCTTAATCCTGATGCGATTATC.....GGGCT 265

351 GlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArgMetProThrTr 367

266 CCACCATTAATCTTCAGATTTCACCGCTGGGGTGGTGTCCACACTT 315

367 pValIleSerProTyrSerLysGlyTyrIleGluGlnTyrGlyThrA 384

316 CTGATTTCCACATGGATGACAGCGCCACTGTGATCCATGACCGAAG 365

384 spProValThrGlyLysProAlaProTyrSerAlaThrSerValLeuLys 400

366 GTCCAACTACGTCTP.....TCACAATTTGAGCATTTCTTCCATACGACA 409

401 ThrLeuGlyTyrLeuTrpAspIleGluAsp.....PheThrProArgVa 415

410 ACTGTGAAAAAGCTTTTAACTCTGAATCAAGTTTCTTAACCAAGCGAGA 459

415 lAlaHisSerProSerPheAspHisLeu.....IleGlyThrThrLeuA 430

460 TGCATGGCTGGGACATTTGAGAACTATTTCCGCTTGCCTAAGACTCTC 509

430 rgGluAspAlaPro 434

510 GTGATGATTGTCCA 523

seq\_name: gb\_est78:BF004644

## seq\_documentation\_block:

LOCUS BF004644 514 bp mRNA EST 06-OCT-2000

DEFINITION EST433142 KV1 Medicago truncatula cDNA clone pkV1-1856, mRNA sequence.

ACCESSION BF004644

VERSION BF004644.1 GI:10704919

KEYWORDS EST.

SOURCE Barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE 1 (bases 1 to 514)

AUTHORS VandenBosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.

TITLE ESTs from roots of Medicago truncatula 24 hours after inoculation with Sinorhizobium meliloti

JOURNAL Unpublished (1999)

COMMENT Contact: VandenBosch K  
Department of Biology  
Texas A&M University  
College Station, TX 77843-3258, USA  
Tel: 409 845 7707  
Fax: 409 845 2891  
Email: kate@mail.bio.tamu.edu

More information is available at: <http://chrysis.tamu.edu/medicago>

Seq primer: SKmod (CTA gAA CTA gTg gAT CC).

## FEATURES

source

1..514

/organism="Medicago truncatula"

/cultivar="genotype A17"

/db\_xref="taxon:3880"

/clone="pkV1-1856"

/clone\_lib="KV1"

/tissue\_type="Seedling roots"

/dev\_stage="24 hours post-inoculation with Sinorhizobium meliloti"

/lab\_host="E. coli strain XL0LR"

/note="vector: pBluescript SK -; Site1: EcoRI; Site2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA







/cultivar="genotype A17"

/db\_xref="taxon:3880"

/clone\_lib="KV1"

/clone\_lib="KV1"

/dev\_stage="24 hours post-inoculation with Sinorhizobium

melliloti"

/lab\_host="E. coli strain XL0LR"

/note="Vector: pBluescript SK -; Site:1: EcoRI; Site:2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the UniZap XR vector from

Stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-Zap phage using Ex-assist

helper phage and propagated in XL0LR cells."

156 a 101 c 125 g 170 t

BASE COUNT  
ORIGIN

alignment\_scores:

Quality: 242.50 Length: 170

Ratio: 2.205 Gaps: 5

Percent Similarity: 64.706 Percent Identity: 34.706

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x BF004645

Align seg 1/1 to: BF004645 from: 1 to: 552

258 PhePheGlnAspAlaTyrLeuGluValLeuProLysPheSerThrValAs 274

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

62 TTCAAGAAAGATCAAGAAATGGGAGCTTCCACCTTTCACAGTGATTGA 111

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

274 nPro.Ser.....CysCysGlyThrAsnThrAsnSerMetHisPro 287

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

112 GCCTAGGTACTTTGACTTGACGGGTTGGCGGCAAAATGATGATCACCCA 161

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

288 ThrGlyAsnValSerTyrGlyGluValPheValLysGlnLeuTyrAspAl 304

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

162 TCTCATGATGTTGCAATGGCAAAATGTTGGTTAAGGAGATTATGAGAC 211

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

304 AtleArgGlnGlyProGlnTrpAspLysThrLeuLeuPheLeuThrTyrA 321

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

212 TCTAAGACCAAGTCCACAGTGGATGAACTTTTGGTCATCATATG 261

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

321 spGluThrGlyGlyPheTyrAspHisValProProLeuAlaValArg 337

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

262 ATGACGATGGTGGATTTTGTGATCATGTGAAGACTCCTTTTGT 305

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

338 ProAspAsnLeuThrTyrThrGluThrAlaLysAsnGlyGlnLysTyrTh 354

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

306 .....ACATTTCTAGTCAGATGGGAACACAGGACCTGCTCTATT 349

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

354 rleuHisPheAspArgLeuGlyGlyArgMetProThrTrpValIleSerP 371

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

350 CTTTAAAGTTTCATAGGTTAGGTGTTCCGGTGCCTACTATTATGCTCTC 399

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

371 rOTyrSerLysGlyTyrIleGluGlnTyrGlyThrAspProValThr 387

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

400 CTTGGATCAAGAGGGATTTGGTGAAGAGCCCAAGGACCGCTGGC 449

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

388 GlyLysProAlaProTyrSerAlaThrSerValLeuLysThrLeuGlyTy 404

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

450 AAT.....TCCGAGTTTGAGCACTCTTCAATCTTCGCCACCATAGAA 493

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

404 rleuTrpAspIleGluAsp.....PheThrProArgValAlaHisSerP 419

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

494 GATGTTCACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 543

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

419 roSerPhe 421

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

544 GCACATTT 551

seq\_name: gb\_est78:BF010466

seq\_documentation\_block:

LOCUS BF010466 392 bp mRNA EST 06-OCT-2000

DEFINITION NCX1\_084\_C01\_F NCX1 (Nsf Xylem Compression wood Inclined) Pinus

taeda cDNA clone NCX1\_084\_C01 5', mRNA sequence.

ACCESSION BF010466

VERSION BF010466.1 GI:10710740

KEYWORDS EST.

SOURCE loblolly pine.

ORGANISM Pinus taeda

REFERENCE 1 (bases 1 to 392)

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

Authors: Sederoff,R.

Journal: Molecular Basis of Wood Formation in the Pine Megagenome

Unpublished (2000)

Contact: Johnson, Arthur

North Carolina State University

Tel: 919 515 7800

Fax: 919 515 7801

Email: ajohnson@unity.ncsu.edu

Seq primer: T3.

Location/Qualifiers

1..392

/organism="Pinus taeda"

/strain="Coastal plain loblolly pine from North Carolina"

/db\_xref="taxon:3352"

/clone\_lib="NCX1\_084\_C01"

/tissue\_type="Xylem"

/cell\_type="Compression"

/dev\_stage="Juvenile"

/lab\_host="XLI-Blue"

/Note="Vector: Bluescript SK; Site:1: Eco RI; Site:2: XhoI

; The library is from early (spring) wood, taken from

three six-year old trees (three different genotypes), in

the juvenile phase. These trees were induced to form

compression wood by bending to a 45 degree angle and tying

them to the ground. Differentiating xylem was harvested

from the bottoms of the inclined stems, and a mixture of

all three genotypes was used for the library. oligo-dr

primed cDNA was directionally cloned into the EcoRI-XhoI

Bluescript SK vector arms. NOTE: The sequences contain a

'cDNA adapter' between the EcoRI site and the start of the

EST. The adapter sequence is 'AATTCGGCAGG'."

BASE COUNT 104 a 93 c 95 g 100 t

ORIGIN

alignment\_scores:

Quality: 241.00 Length: 148

Ratio: 2.648 Gaps: 5

Percent Similarity: 61.486 Percent Identity: 37.162

alignment\_block:

US-09-426-072-2\_COPY\_20\_464 x BF010466

Align seg 1/1 to: BF010466 from: 1 to: 392

251 AsnValProValGluAsnPhePheGlnAspAlaTyrLeuGlyValLe 267

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

10 AATTATGGTGTGATCGAGCAGCTTACTATGATCTCAAGCTTGGC..... 54

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

267 uProLysPheSerTyrIleAsnProSerCysGlyThrAsnThrAsnS 284

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

55 .....CCTGGA.....AATG 64

284 erMetHisProThrGlyAsnValSerTyrGlyGluValPheValLysGln 300

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

65 ATGACCATCGTCCCAAGATGTTGGTGAAGGCGAGAGTTTGTGAAGAG 114

||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

301 lleTyrAspAlaIleArgGlnGlyProGlnTrpAspLysThrLeuLeuPh 317

```

115 GTGATGAGGTGCTGAGATCGATCCCGAGTGAACGAGATGTTGTTTCAT 164
116 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
317 eilethrtvraspGluThrGlyGlyPheTyAspHisValProProL 334
318 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
165 TATCATTACATGAGTACGACGGGGTTCAGATCAGTATGCGCCAGCGCTG 214
334 eu.....AlaValArgProAspAsnLeuThrTyThrGluThrAlaLys 348
215 TGAATAATGTCCTCCAGATGATATCGTTGGATCAGAG..... 255
349 AsnGlyGlnLysThrLeuHisPheAspArgLeuGlyGlyValArgMetPr 365
256 .....CCTATTACTTCAAAATCGACAGACTGGCGTCCGTTGATACC 296
365 oThrTrpValIleSerProTyThrSerLysGlyTyThrGluGlnTyrg 382
297 GACTATCATGATTTCCTCTGGATCAACAGGGAACAGTG..... 336
382 lyThrAspProValThrGlyLysProAlaProTyThrSerAlaThr 396
337 .....GTACATGGCCCAATGGACCAATATCTTCTCTCC 369

```

seq\_name: gb\_est9:AA556789

```

seq_documentation_block:
LOCUS AA556789 661 bp mRNA EST 28-AUG-1998
DEFINITION 631 Loblolly pine NA Pinus taeda cDNA clone 2NA9B, mRNA sequence.
ACCESSION AA556789
VERSION AA556789.1 GI:3365803
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda
REFERENCE 1 (bases 1 to 661)
AUTHORS Allona, I., Quinn, M., Shoop, E., Swope, K., St. Cyr, S., Carlis, J.,
Riedl, J., Retzel, E., Campbell, M. M., Sederoff, R., and Whetten, R. W.
TITLE Analysis of xylem formation in pine by cDNA sequencing
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9693-9698 (1998)
MEDLINE 98356220
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: T3.

```

```

FEATURES
Source
1..661
/organism="Pinus taeda"
/strain="Coastal plain loblolly pine from North Carolina"
/db_xref="taxon:3352"
/clone="2NA9B"
/tissue_type="xylem"
/lab_host="SOLR"
/note="Vector: lambda-ZAP; Site_1: EcoRI; Site_2: XhoI;
The result of subtraction of library N with library C.
Immature xylem from the side of inclined stems of
differentiating wood was subtracted with immature xylem
from underside of inclined stems of differentiating
compression wood. A mixture of four genotypes were used.
oligo-dT primed cDNA was directionally cloned into the
EcoRI-XhoI lambda-ZAP vector arms"
BASE COUNT 154 a 163 c 158 g 160 t 26 others
ORIGIN

```

alignment\_scores: Quality: 236.50 Length: 174

```

Ratio: 2.112 Gaps: 6
Percent Similarity: 64.368 Percent Identity: 32.759
alignment_block:
US-09-426-072-2_COPY_20_464 x AA556789 ..
Align seg 1/1 to: AA556789 from: 1 to: 661
265 GlyValLeuProLysPheSerTyThrIleAsnPro.....SerCysCy 278
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 GGGGGGCTGCCAAATGCTGGTGCATAGANCCCGCTACTTCCATCTCGC 50
278 sGlyThrAsnThrAsnSerMetHisProThrGlyAsnValSerTyGlyG 295
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 GATTTCNCTCNAAACGATGACCATCTTCGATGATGATTTCCACGGGC 100
295 luValPheValLysGlnIleTyAspAlaIleArgGlnGlyProGlnTrp 311
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 AGAATCTGCTGAGGAGGTTTACGAAGCGCTGANAGCAANCCCGCAGTGG 150
312 AspLysThrLeuLeuPheIleThrTyAspGluThrGlyGlyPheTyAs 328
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 ACCAAACCTGCTGCTCTATTACATACGATGACACGGCGGTTTCTAGA 200
328 phiSValProProLeu.....AlaValArgProAspAsnLeuThr 343
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 TCATGTGCCACCGCAGTTCGCAATGTCGCCAGCCCGCAGCGCATTTGG 249
343 yrThrGluThrAlaLysAsnGlyGlnLysTyThrLeuHisPheAspArg 359
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
250 .....GGATCCCTCCCTTTTTCATTTTTCATTTTTCATTTTTCAGAGA 282
360 LeuGlyGlyArgMetProThrTrpValIleSerProTyThrSerLysLysG 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
283 TTAGCGTTCGCTGCTCCACATATATGTTGCTCCCTTGGATCAACAAAG 332
376 yTyThrGluGlnThrGlyThrAspProValThrGlyLysProAlaProT 393
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
333 AACNTGGTGGACCGCCCTAGCGGGCCA.....TCGCCAGCTCAGANT 376
393 yrSerAlaThrSerValLeuLysThrLeuGlyTyThrLeuTrpAspIleGlu 409
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
377 TCGAGCATTCATGATTCCTCGAATCATATCAAGAAATGTTCAATCTCTCT 426
410 Asp.....PheThrProArgValAlaHisSerProSerPheAspHisLe 424
427 GCCAATTTTCTCACTACAGANATGCCCTGGCTGGAACTTTCGAAAGGCT 476
424 ulleGlyThrThrLeuArgGlu 431
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
477 NGTG.....AGCTTGGCGGAT 492
seq_name: gb_est42:AW266012

```

```

seq_documentation_block:
LOCUS AW266012 601 bp mRNA EST 29-DEC-1999
DEFINITION L30-2674T3 Ice plant Lambda Uni-zap XR expression library, 30 hours
NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-2674
5', mRNA sequence.
ACCESSION AW266012
VERSION AW266012.1 GI:6644699
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;
Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 601)
AUTHORS Cushman, J.C.
TITLE An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
JOURNAL Unpublished (1997)
COMMENT Contact: Cushman JC

```



```
202 CCACACCCCATACATCAATGTCACAGCCAGATGGAGACACCGGTCGG 251
347 laLysAsnGlyGlnLysTyrThrLeuHisPheAspArgLeuGlyGlyArg 363
   ||::: ||| ||||| ||||| |||
252 CTCCTTCCTTCTCAAG.....TTCGATAGGCCTCGGTGTTCCA 289
364 MetProThrTrpValIleSerProTyrSerLysGlyTyrIleGluG1 380
   ::||| ||::: ||||| :: ||||| ::
290 GTCCGACGATCATGGTCTCTCCTGGATCAAGAAAGGCACTGTG..... 334
380 nTyrGlyThrAspProValThrGlyLysProAlapro..... 392
   ::::: ||| ::||
335 .....ATTAGTGGTCCGAGGGTCCTTCCCGAACTCTG 368
393 ..TyrSerAlaThrSerValLeuLysThrLeuGlyTyrLeuTrpAspIle 408
   ::::: ||||| ::||| ::::
369 AATTCGAGCACTCATCGATCCCTGCAACGATAAAAAAATGTCAATCTA 418
409 GluAsp.....PheThrProArgValAlaHisSerProSerPheAspHi 423
   ::::: |||| ||| ::::: ||||| ::::
419 TCTTCCAACCTCTTGACTCATAGAGATGCTTGGCGCGCACCTTCGAACG 468
423 sLeuIle 425
   :::::
469 CGTTGTC 475
```



```

; APPLICANT: Klionsky, Daniel
; APPLICANT: Holzer, Helmut
; APPLICANT: Destruelle, Monica
; TITLE OF INVENTION: STRESS TOLERANT YEAST MUTANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,714
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard Aton
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-61036/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 210 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-10782-2

```

```

Query Match          6.5%: Score 86.5; DB 1; Length 1041;
Best Local Similarity 21.2%: Pred. No. 2.8; Indels 77; Gaps 16;
Matches 61; Conservative 45; Mismatches 105;

```

```

QY 3 VTSEY---TSVRAPFGYKPGSKESIEENLKDKNVIVWLILENRS-----FQNLGG 51
Db 446 VDYEYHINVDVAMASFSIYLNHGEHIDD--EYLYRTTWPITKNAQAFFYAYKYNSLGL 503
QY 52 VRRGLDNP-----INNPGFCNYKNASDPSSGKYCTQAKDYDSVNDPDHVSVCNNLEF 105
Db 504 YETYNLTDPDEFANHNGAF---TNAGIKTLKWKATDIGNHLGEVVDPKWSEISKDI-- 558
QY 106 YGTVTNNGAISKGVADQSGFLNQLNDYKPLAPEATQV--MGYYTEEEVPTLVDL 163
Db 559 ---YIPR---SSNITLEYSG-----MNSVVEIKQADVTLMLVPLGYINDESI--LNNA 604
QY 164 VDEFTTNSWFCV--PGPTNPNRLCALAGTAAGH-----K 198
Db 605 IKDLYYYSERQASGAMPYVFVAAAAG--LLNHGSSSQSYLYKSVLPYLRAPFAQFSEQ 663
QY 199 NDDDFLNYGFSKSI--FEANEK-----GVSWLNYDGTNGEFPDS 238
Db 664 SDDNFLTNGLTQAPFFLTANGFLOSILFGUTGIRY---SYEVDPT 708

```

```

RESULT 3
PCT-US96-10782-2
; Sequence 2, Application PC/TUS9610782
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University
; APPLICANT: of California
; TITLE OF INVENTION: STRESS TOLERANT YEAST MUTANTS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBBINS, BERLINER & CARSON

```

```

; STREET: 201 N. Figueroa Street, 5th Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90012-2628
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10782
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Berliner, Robert
; REGISTRATION NUMBER: 20,121
; REFERENCE/DOCKET NUMBER: 5555-400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 977-1001
; TELEFAX: (213) 977-1003
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-10782-2

```

```

Query Match          6.5%: Score 86.5; DB 4; Length 1041;
Best Local Similarity 21.2%: Pred. No. 2.8; Indels 77; Gaps 16;
Matches 61; Conservative 45; Mismatches 105;

```

```

QY 3 VTSEY---TSVRAPFGYKPGSKESIEENLKDKNVIVWLILENRS-----FQNLGG 51
Db 446 VDYEYHINVDVAMASFSIYLNHGEHIDD--EYLYRTTWPITKNAQAFFYAYKYNSLGL 503
QY 52 VRRGLDNP-----INNPGFCNYKNASDPSSGKYCTQAKDYDSVNDPDHVSVCNNLEF 105
Db 504 YETYNLTDPDEFANHNGAF---TNAGIKTLKWKATDIGNHLGEVVDPKWSEISKDI-- 558
QY 106 YGTVTNNGAISKGVADQSGFLNQLNDYKPLAPEATQV--MGYYTEEEVPTLVDL 163
Db 559 ---YIPR---SSNITLEYSG-----MNSVVEIKQADVTLMLVPLGYINDESI--LNNA 604
QY 164 VDEFTTNSWFCV--PGPTNPNRLCALAGTAAGH-----K 198
Db 605 IKDLYYYSERQASGAMPYVFVAAAAG--LLNHGSSSQSYLYKSVLPYLRAPFAQFSEQ 663
QY 199 NDDDFLNYGFSKSI--FEANEK-----GVSWLNYDGTNGEFPDS 238
Db 664 SDDNFLTNGLTQAPFFLTANGFLOSILFGUTGIRY---SYEVDPT 708

```

```

RESULT 4
US-08-470-350B-2
; Sequence 2, Application US/08470350B
; Patent No. 5684126
; GENERAL INFORMATION:
; APPLICANT: Li, Xiao
; APPLICANT: Snyder, Solomon H
; TITLE OF INVENTION: Ebnerin: A Secreted von Ebner's Gland
; TITLE OF INVENTION: Protein Associated with Taste Buds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001

```





```

RESULT      8
PCT-US96/07528-2
; Sequence 2, Application PC/TUS9607528
; GENERAL INFORMATION:
; APPLICANT: The Jackson Laboratory
; TITLE OF INVENTION: CLINICAL DISORDERS ASSOCIATED WITH
; TITLE OF INVENTION: CARBOXYPEPTIDASE E MUTATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07528
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,262
; FILING DATE: 26-MAY-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: JL-9501 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 2:

```

SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-07528-2

Query Match 6.3% Score 83.5; DB 4; Length 435;  
Best Local Similarity 23.1%; Pred No. 1.5;  
Matches 63; Conservative 26; Mismatches 67; Indels 117; Gaps 16;  
QY 33 ENIWIILNRFSDNIGGVRRQGLDNPNNPCFNKNSDPSS-----GKXCTQAKD 86  
DB 99 ETIVNLIHSTRI-HIMPSLNDGFE-----KAASQPGELKDFVGRSNAQCID 145  
QY 87 YDSVFNDPHSHTGNLEFYGTTPNNGAIASGVKVVADQSGFLNAQLNDYPKLAPEATR 146  
DB 146 LNRNFDLRIYVVEKE---GGPNHLLKNMKIYDQN-----TKLAPE--TK 189  
QY 147 QVQCYTTEEVPTLV-----DLV-----DE-----FTTF 170  
DB 190 AVHTWM--DIPFVLSANLHGGDLVANYPYDETRGSAHEYSSPDDAIFQSLARAYSSF 247  
QY 171 NSWFSCVPGTNPNR-LCALAGTAAGHKNDLFLNYGSSKSIFFAANEKGVSW----- 224  
DB 248 N-----PAMSDNRPPCR-----KNDD-----SSFVDGTTNGGAWYSPG 283  
QY 225 -----LNYDGTN-----GEFEPDSLFFTY 243  
DB 284 GMQDFNYLSSNCFEITVELSCERFPPEETLKY 316

RESULT 9  
US-08-928-361B-11  
; Sequence 11, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Catolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,361B  
; FILING DATE: 12-SEP-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,062  
; FILING DATE: 13-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Verny, Hana  
; REGISTRATION NUMBER: 30,518  
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-1677  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1042 amino acids

TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-928-361B-11

Query Match 6.2% Score 82.5; DB 3; Length 1042;  
Best Local Similarity 20.8%; Pred No. 7.3;  
Matches 50; Conservative 21; Mismatches 66; Indels 103; Gaps 9;  
QY 59 NPINNGPCFNKNSADPSSKCYC-----TOAKDYDSVFN----- 92  
DB 824 DPISYLPFAKNGELIDPISGYFSGSIAGFISGKAGSQSKSSDESNGPIDPSTNNPYDPK 883  
QY 93 -----DP-----DSHVTGNLEFYGTYPNNGAIASGVKVVADQSGFLNAQLNDYPKLAP 141  
DB 884 TGKLIDPESGIAIDNSYGVFAIVPGTAAPKGGVIVPESVAAE-----AAK 929  
QY 142 EETATQVMGYTTEEVPTLVLDLDEFTTNSWFCVPGTNPNRLC--ALAGTAAG----- 195  
DB 930 KYFAANVEGEVEEVP-----PPPESSNIAIQAAGGASAAVGLV 970  
QY 196 -----HGKNDLFLNYGSSKSIFFAANEKGV-----SMLNYDGTN 231  
DB 971 AAVGANVASRNRQEGEDDD---YQMDLQNNKKKKRYMKQOMKLLQLLSVHHSGTN 1027

RESULT 10  
US-08-928-361B-30  
; Sequence 30, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Catolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,361B  
; FILING DATE: 12-SEP-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,062  
; FILING DATE: 13-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Verny, Hana  
; REGISTRATION NUMBER: 30,518  
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-1677  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1043 amino acids  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-928-361B-30

Query Match 6.2%; Score 82.5; DB 3; Length 1043;  
 Best Local Similarity 20.8%; Pred. No. 7.3;  
 Matches 50; Conservative 21; Mismatches 66; Indels 103; Gaps 9;

Qy 59 NPINNGPCNYKNASDPSSGKYC-----TQAKDYDSVFN----- 92  
 Db 825 DPISVLPFAKNGELIDPISGKYFSGIAGFISGKAGSQSKSDSGNPIDPSTNMPYDPK 884  
 Qy 93 -----DP-----DHSVTGNLLEFYGYTPNNGAIASGKVADQSGFLNAQLNDYPKLAP 141  
 Db 885 GKKLIDPESGIAIDNSVSGVFATPGTAAPKKGVIPEVAE-----AAK 930  
 Qy 142 EATRQVMGYTTEEVPTLVLDVDEFTTNSWFSVPGPTNPNRLC--ALAGTAAG---- 195  
 Db 931 KYFAANVEGEGEVEVP-----PPPESSSNIAIQAAGGASAAVGLV 971  
 Qy 196 -----HGKNDLDFLNYGSISSKSIPEAAANEKV-----SWLNYDGTN 231  
 Db 972 AAVGAWYASNRQEGEDDD---YQMDLKQNMKKRKRVMKQMKLLQLSVIHSGTN 1028

RESULT 11  
 US-08-700-651-5  
 ; Sequence 5, Application US/08700651B  
 ; Patent No. 6015882  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PETERSEN, CAROLYN  
 ; APPLICANT: LEECH, JAMES  
 ; APPLICANT: NELSON, RICHARD, C.  
 ; APPLICANT: GUT, JIRI  
 ; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
 ; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
 ; TITLE OF INVENTION: INFECTIONS  
 ; FILE REFERENCE: 480.19-4(HV)  
 ; CURRENT APPLICATION NUMBER: US/08/700,651B  
 ; EARLIER FILING DATE: 1997-08-14  
 ; EARLIER APPLICATION NUMBER: 08/415,751  
 ; EARLIER FILING DATE: 1995-04-03  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 5  
 ; LENGTH: 1721  
 ; TYPE: PRT  
 ; ORGANISM: Cryptosporidium parvum  
 US-08-700-651-5

Query Match 6.2%; Score 82.5; DB 3; Length 1721;  
 Best Local Similarity 20.8%; Pred. No. 16;  
 Matches 50; Conservative 21; Mismatches 66; Indels 103; Gaps 9;

Qy 59 NPINNGPCNYKNASDPSSGKYC-----TQAKDYDSVFN----- 92  
 Db 1503 DPISVLPFAKNGELIDPISGKYFSGIAGFISGKAGSQSKSDSGNPIDPSTNMPYDPK 1562  
 Qy 93 -----DP-----DHSVTGNLLEFYGYTPNNGAIASGKVADQSGFLNAQLNDYPKLAP 141  
 Db 1563 GKKLIDPESGIAIDNSVSGVFATPGTAAPKKGVIPEVAE-----AAK 1608  
 Qy 142 EATRQVMGYTTEEVPTLVLDVDEFTTNSWFSVPGPTNPNRLC--ALAGTAAG---- 195  
 Db 1609 KYFAANVEGEGEVEVP-----PPPESSSNIAIQAAGGASAAVGLV 1649  
 Qy 196 -----HGKNDLDFLNYGSISSKSIPEAAANEKV-----SWLNYDGTN 231  
 Db 1650 AAVGAWYASNRQEGEDDD---YQMDLKQNMKKRKRVMKQMKLLQLSVIHSGTN 1706

RESULT 12  
 US-08-928-361B-6  
 ; Sequence 6, Application US/08928361B

Patent No. 6071518  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Petersen, Carolyn  
 ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
 ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
 ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
 ; TITLE OF INVENTION: SPECIES INFECTIONS  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
 ; STREET: 385 Sherman Avenue, Suite 6  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94306-1840  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/928,361B  
 ; FILING DATE: 12-SEP-1997  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/026,062  
 ; FILING DATE: 13-SEP-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Verny, Hana  
 ; REGISTRATION NUMBER: 30,518  
 ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-324-1677  
 ; TELEFAX: 650-324-1678  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1721 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-928-361B-6

Query Match 6.2%; Score 82.5; DB 3; Length 1721;  
 Best Local Similarity 20.8%; Pred. No. 16;  
 Matches 50; Conservative 21; Mismatches 66; Indels 103; Gaps 9;

Qy 59 NPINNGPCNYKNASDPSSGKYC-----TQAKDYDSVFN----- 92  
 Db 1503 DPISVLPFAKNGELIDPISGKYFSGIAGFISGKAGSQSKSDSGNPIDPSTNMPYDPK 1562  
 Qy 93 -----DP-----DHSVTGNLLEFYGYTPNNGAIASGKVADQSGFLNAQLNDYPKLAP 141  
 Db 1563 GKKLIDPESGIAIDNSVSGVFATPGTAAPKKGVIPEVAE-----AAK 1608  
 Qy 142 EATRQVMGYTTEEVPTLVLDVDEFTTNSWFSVPGPTNPNRLC--ALAGTAAG---- 195  
 Db 1609 KYFAANVEGEGEVEVP-----PPPESSSNIAIQAAGGASAAVGLV 1649  
 Qy 196 -----HGKNDLDFLNYGSISSKSIPEAAANEKV-----SWLNYDGTN 231  
 Db 1650 AAVGAWYASNRQEGEDDD---YQMDLKQNMKKRKRVMKQMKLLQLSVIHSGTN 1706

RESULT 13  
 US-08-928-361B-5  
 ; Sequence 5, Application US/08928361B  
 ; Patent No. 6071518  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Petersen, Carolyn  
 ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
 ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

```

; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1837 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-5

```

```

Query Match      6.2%; Score 82.5; DB 3; Length 1837;
Best Local Similarity 20.8%; Pred. No. 18;
Matches 50; Conservative 21; Mismatches 66; Indels 103; Gaps 9;

QY 59 NPTNNGPFCNYKNASDPSSGKYC-----TQAKDYDSVN----- 92
Db 1619 DPISYLPFAKNGELIDPSIGKYGSGIAGFTSGKAGSOSKSSDESNDPSTNPPYDPK 1678
QY 93 -----DP-----DHSVTGNLLEFYGTYPNNGAIGASKYVADQSGFLNAQLNDYPKLP 141
Db 1679 TGLIDPESGIAIDNSVSGVFATVPCTAAPKKGVPESVAE-----AAK 1724
QY 142 EEAATROVMGYTTEEVPTLVLDVDEFTTNSWFCVCPNPNRLC--ALACTAAG---- 195
Db 1725 KYFAANVEGEGEVPP-----PPPESSNIAIQAGGASAGVGLV 1765
QY 196 -----HGKNDPDLFLYGISSKSIFFAANEKGV-----SWLNVDGNT 231
Db 1766 AAVGAWYASRNREQEDDDDD---YQMDLQNMKKRKRVMKQOMKLLQLLSVLIHSGTN 1822

```

```

RESULT 14
US-08-111-939-14
; Sequence 14, Application US/08111939
; Patent No. 5460951
; GENERAL INFORMATION:
; APPLICANT: Kawai, Shinji
; APPLICANT: Takeshita, Sunao
; APPLICANT: Okazaki, Makoto
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
; TITLE OF INVENTION: Protein and Process for its Production
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/111,939
; FILING DATE: 26-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324033/92
; FILING DATE: 03-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 230029/92
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1321-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4000
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-111-939-14

```

```

Query Match      6.2%; Score 82; DB 1; Length 435;
Best Local Similarity 23.8%; Pred. No. 2.2;
Matches 57; Conservative 25; Mismatches 71; Indels 86; Gaps 13;

QY 33 ENIVWLEILNFRSFDNLLGGVPROGLDNPINNGPFCNYKNASDPSS-----GKYCTQAKD 86
Db 99 ETIVNLHSTRI--HIMPSLNPDGFE-----KAASQPGELKDMFVGRSNAQCID 145
QY 87 YDSVFNDPDHRSVTGNLLEFYGTYPNNGAIGASKYVADQSGFLNAQLNDYPKLAPEEATR 146
Db 146 LNRNFPDLDRIVVYVNEKE---GPNHLLKLNKKIVDQNS-----KLAPE--TK 189
QY 147 QVMGYTTEEVPTLV-----DLVDEFT-----TFNSWFSCV----- 177
Db 190 AVIHWIM--DIFVLUSANLHGGDLVANYPYDETRSGTAHEYSSCPDDAIFQSLARAYSSF 247
QY 178 -PGPTNPNR-LCALAGTAAGKGNDDDFLNLGYGISSKSIFFAANEKGVSWLNDGTNGE 234
Db 248 NPVMSDPNPPCR-----KNDDD-----SSFVDGTNGGANYSVPGGMQDF 288

```

```

RESULT 15
US-08-111-939-15
; Sequence 15, Application US/08111939
; Patent No. 5460951
; GENERAL INFORMATION:
; APPLICANT: Kawai, Shinji
; APPLICANT: Takeshita, Sunao
; APPLICANT: Okazaki, Makoto
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
; TITLE OF INVENTION: Protein and Process for its Production
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

```

Search completed: February 26, 2001, 16:20:38  
Job time: 43368 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2001, 16:20:39 ; Search time 71.55 Seconds  
(without alignments)  
760.090 Million cell updates/sec

Title: US-09-426-072-2

Perfect score: 2501

Sequence: 1 MKSTALLTGLGLASGLAS.....GTTLEDAPALKTPTFTSV 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL15:

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp Vertebrate:\*
- 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	18.7	523	10 Q9SR06	Q9srq6 arabidopsis
2	458	18.3	538	10 Q9SR07	Q9srq7 arabidopsis
3	453.5	18.1	533	10 Q9LML5	Q9lml5 arabidopsis
4	437	17.5	521	10 Q9S816	Q9s816 arabidopsis
5	436	17.4	462	10 Q9SMN8	Q9smn8 arabidopsis
6	435.5	17.4	514	10 Q81020	Q81020 arabidopsis
7	324.5	13.0	508	2 P95245	P95245 mycobacteri
8	318	12.7	700	2 Q9RGS8	Q9rgs8 burkholderi
9	308	12.3	514	2 Q9XB13	Q9xb13 mycobacteri
10	288	11.5	512	2 P95246	P95246 mycobacteri
11	286	11.4	520	2 Q50560	Q50560 mycobacteri
12	271	10.8	521	2 Q50561	Q50561 mycobacteri
13	206	8.2	577	2 Q9L5C5	Q9l5c5 burkholderi
14	203.5	8.1	514	2 Q47936	Q47936 francisella
15	202.5	8.1	501	2 Q9XAN9	Q9xan9 streptomyce
16	196	7.8	577	2 Q9L5C4	Q9l5c4 burkholderi
17	195	7.8	577	2 Q9KH04	Q9kh04 mycobacteri
18	158	6.3	280	2 Q06792	Q06792 mycobacteri
19	128	5.1	221	2 Q9XAH3	Q9xah3 streptomyce

20	116	4.6	2178	2	Q46149	Q46149 clostridium
21	112	4.5	428	10	Q04851	Q04851 brassica na
22	112	4.5	428	10	Q9M429	Q9m429 brassica na
23	111	4.4	428	10	Q42624	Q42624 brassica na
24	111	4.4	622	10	Q9S9K5	Q9s9k5 arabidopsis
25	111	4.4	629	10	P93217	P93217 lycopersico
26	109	4.4	299	2	O53361	O53361 mycobacteri
27	109	4.4	430	10	Q9SEX6	Q9sex6 canavalia l
28	109	4.4	440	10	Q9SNJ3	Q9snj3 oryza sativ
29	108.5	4.3	492	2	O52155	O52155 burkholderi
30	108.5	4.3	626	10	O64599	O64599 arabidopsis
31	107.5	4.3	492	2	Q9ADX7	Q9adx7 burkholderi
32	107.5	4.3	1766	12	Q9J599	Q9j599 fowlpox vir
33	107	4.3	724	12	Q9Y1J1	Q9y1j1 adeno-assoc
34	106	4.2	622	10	P92982	P92982 arabidopsis
35	105.5	4.2	962	5	O9G164	O9g164 plasmodium
36	105	4.2	305	2	Q9KCH5	Q9kch5 bacillus ha
37	105	4.2	938	12	Q9W8S1	Q9w8s1 porcine ade
38	105	4.2	939	12	Q9YTR8	Q9ytr8 porcine ade
39	105	4.2	2229	5	Q19853	Q19853 caenorhabdl
40	104.5	4.2	1687	2	Q9R9B7	Q9r9b7 porphyromon
41	104.5	4.2	1704	2	O51816	O51816 porphyromon
42	104.5	4.2	4589	11	Q9WU10	Q9wu10 rattus norv
43	104	4.2	428	8	Q9XQ94	Q9xq94 medicago sa
44	104	4.2	432	10	Q9SPJ1	Q9spj1 juglans nig
45	103.5	4.1	720	3	O14367	O14367 schizosacch

## ALIGNMENTS

RESULT 1

Q9SR06 PRELIMINARY; PRT; 523 AA.

AC Q9SR06

DT 01-MAY-2000 (TremBLrel. 13, Created)

DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)

DE T21P5.6 PROTEIN.

GN T21P5.6.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Rosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,

RA Ronning C.M., Fujii C.Y., Utterback T.R., Barnstead M.E.,

RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;

RT "Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence.";

RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC009895; AAF01583.1; -

SQ SEQUENCE 523 AA; 59097 MW; 31D10D4985C41533 CRC64;

Query Match 18.7%; Score 468; DB 10; Length 523;  
Best Local Similarity 30.8%; Pred. No. 6.2e-28;  
Matches 132; Conservative 74; Mismatches 167; Indels 56; Gaps 16;

Qy	51	VENIVMLLENLNSFDNIIIGYVRRG--LDNPINNGFCNYKNASDPSSGKYQAKYDS	108
Db	16	IKTIVLVQNKNSFDHMLGNFKELNPEDVGVSESEPRSNPLUSTDPNSAQIFF-QRESQN	74
Qy	109	VFNDPDHVSNGNNLEFYGTYPNNGATASGVVADQSGFLNAQLNDYKPLAPEAT----	164
Db	75	IDDPGHS-----FQAIYEQ-----VFGKPFSESPYDPKMGVQNA-EAITKMS	121
Qy	165	-RQVNGVYEEVEEYPLVDVDEFTTFNSWFCVPGTNPENLALACTAAGHCKNDDDL	223
Db	122	EKVYMQGFPPEKPLVFKELVQFAVCDRMFSLSPSTQPNRLYVHAATSGAFSNDTNTL	181





Db 175 PNRVYVSHSATHGSSNKKKDLVKGFPQKTFIDSLDENGLSP----GIYYONIPATFFPK 230  
 QY 262 YVNOTSRNVPVEN----FQDAYLGLVLPKFSYINPSCCGTN---TNSMHPGTGNVSKGE 314  
 Db 231 SLRRL--KHLVKFHSYALKLDAKGLKLPNSVVEQRYFDIDLFPANDDHPSHDVAAGQ 288  
 QY 315 VEVQIYDAIROGPOWDTLLFTYDEGGFYDHPVPL--AVRPDLNLTETAKNGOKY 372  
 Db 289 REVKEVETLRSSPQWKEALLIYDEHGGFYDHPVTPVKGVPMNDGII-----GPDPPF 342  
 QY 373 TLHEDRGLGRMPTWVISPYSKGYI--EYGTGDPVTGKPAPYSA---TSVLKTLGLYLDI 427  
 Db 343 YEGFDRLGVVRPFLISSMIEKGTVIHEPEG-----PTPHSQFESSIPATVKKLNL 395  
 QY 428 ED--FTPRVAHSPFDHL--IGTLTREDAP 453  
 Db 396 KSHFLTKRDWAGTFEKYFRINDSPRODCP 425

RESULT 4

Q9S816 PRELIMINARY; PRT; 521 AA.  
 ID Q9S816  
 AC Q9S816;  
 DT 01-MAY-2000 (T:EMBLrel. 13, Created)  
 DT 01-MAY-2000 (T:EMBLrel. 13, Last sequence update)  
 DE PUTATIVE PHOSPHOLIPASE.  
 GN T12J13.18 OR T21P5.4.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,  
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.,  
 RT "Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AAC09327; AAF03477.1;  
 DR EMBL; AAC09895; AAF01581.1;  
 SQ SEQUENCE 521 AA; 59071 MW; 1A39ACEF6807B8EF CRC64;

Query Match 17.5%; Score 437; DB 10; Length 521;  
 Best Local Similarity 30.0%; Pred. No. 1.5e-25;  
 Matches 133; Conservative 73; Mismatches 161; Indels 76; Gaps 18;  
 QY 51 VENIWLILNRSFDNIGL-----GVRRL-----GLDNPIN--NGPFCNYKNAS 92  
 Db 13 KTIWLVQENRSFQHTLWGFKEINREIDGVKMSQKFNPGFSSDLNHNRYVFGDSQYV 72  
 QY 93 DPSSGKYCTQAKD-YDSVFNPDHSHVTGNLLEFYGTYPNNGATASGVVADQSGFL-NA 150  
 Db 73 DPNFGH---SIRDIYEQVFGPWS-----GHPDPNPGP-----ATWSGFAQNA 113  
 QY 151 QINDYPKLAPEATROVMGYTTEEEVPTLVLDVDEFTTNSWFSVPGPTNPNRCLALAG 210  
 Db 114 ERK-----MKGMSSAVMNGFKPDALPVYKELVQNFALCDRWFAVPGATQPNRLFIHSA 167  
 QY 211 TAAGHGKDDDFLNYGISSKSIPEANNEKGVSNLNYDGTNGEFGEPDSLEFTTVNQ---TS 267  
 Db 168 TSHGTTNNEKLLIEFGFKQTFESLDEAGTF-----GIYQCPPTTLFRLNRLKLYLT 223  
 QY 268 RSNVVPVENFQDAYLGLVLPKFSYI-----NPSCCGTNTNSMHPGTGNVSYGEVEVK 318  
 Db 224 RFHDYGLQ-FKKDCKREGNLPNYYVVEQRYWYDILLNPA-----NDQHPSHVSEGOGLVK 276  
 QY 319 QYDAIROGPOWDTLLFTYDEGGFYDHPVPLAVRPDLNLTETAKNGOKYTLHEDR 378  
 Db 277 EYELARSSPQWKEALLIYDEHGGFYDHPVTPPL-----DCVPNDGILGPPPYFENR 332

QY 379 LGRMPTWVISPYSKGYIEQYTDPTVTKGKPAPYSATSVLKTGLGLYLDIEDF--TPRVAHS 437  
 Db 333 LGRVVPFFISPIWIEGTVHLSNCPYL--MSQYEHSSIPATVKKIFLKADFLTKRDSWA 390  
 QY 438 PSFDHLI-GTTLREDAPIALKTP 459  
 Db 391 GTFESVITRNSPRQDCPETLSNP 413

RESULT 5

Q9SMN8 PRELIMINARY; PRT; 462 AA.  
 ID Q9SMN8  
 AC Q9SMN8;  
 DT 01-MAY-2000 (T:EMBLrel. 13, Created)  
 DT 01-MAY-2000 (T:EMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (T:EMBLrel. 13, Last annotation update)  
 DE HYPOTHETICAL 51.3 KDA PROTEIN.  
 GN T8P19.120.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cholsne N., Robert C., Brottier P., Wincker P., Cattolico L.,  
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X., Queciet F., Salanoubat M.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL13315; CAB62350.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 462 AA; 51284 MW; EE6492B529991095 CRC64;

Query Match 17.4%; Score 436; DB 10; Length 462;  
 Best Local Similarity 32.0%; Pred. No. 1.5e-25;  
 Matches 125; Conservative 61; Mismatches 165; Indels 40; Gaps 13;  
 QY 78 IPINN---GPFCNYKNASDPSSGKYCTQAKDYDSVFNPDHSHVTGNLLEFYGTYPNNGA 134  
 Db 6 NPTINGVTGQECN----PVPNSTQTICFTSDAEFVDPDPGHSFEAVQQVFGS---GFGQ 58  
 QY 135 TASGVVADQSGFLNQLNDYPKLAPEATROVMGYTTEEEVPTLVLDVDEFTTNSWFS 194  
 Db 59 IPS-----MMGFVEQALS-----MPGNLSETVMKGFPRPEAVPVYAEVKEFAVDFRWS 107  
 QY 195 CVPGPTNPNRCLALAGTAGHGKNDLFLNYGISSKSIPEANNEKGVSNLNYDGTNGEFE 254  
 Db 108 SIPGTPONRLFYVSTSHGTSHTSVHKQLAQQYPOKTFIDSLHNDIDF----GIYFQNI 163  
 QY 255 PDSLFYTVNQTSR--SNVVPVE--NFODAYLGLVLPKFSYINP---SCCGTNTNSMHPGTGN 309  
 Db 164 PTLTYRNLRLQKLIYFNLDLAFKDAKGLPSLVIEPRYFDLKLGLPANDDHPSHD 223  
 QY 310 VSYGEVVKIYDAIROGPOWDTLLFTYDEGGFYDHPVPLAVRPDLNLTETAKNG 369  
 Db 224 VANGOKLYEVEALRSSPQWNETLLIYDEHGGFYDHPVTPVIGIPNDGNTGAPGF 283  
 QY 370 QKYLTHEDRGLGRMPTWVISPYSKGYIEQYTDPTVTKGKPAPYSATSVLKTGLYLDIED 429  
 Db 284 FK----FDRGLGVVRPTIMVSPWIKQGVVSEAKGPT--ESSEYEHSSIPATIKKLFNLSS 337  
 QY 430 ---FTPRVAHSPFDHLIG--TTLREDAPIAL 456  
 Db 338 NFLTHRDWAGTFEDVSHLTPTRDCPWL 368

RESULT 6

O81020  
 AC O81020; PRELIMINARY; PRT; 514 AA.  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DE F12C20-9 PROTEIN.  
 GN F12C20-9.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 CC Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Roundley S.D., Ronning C.M., Lin X., Ketchum K.A., Crosby M.L.,  
 RA Bransley R.C., Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R.,  
 RA Adams M.D., Somerville C.R., Venter J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC005168; AAC32238.1; -.  
 SQ SEQUENCE 514 AA; 57681 MW; 0575269A64F1C2CE CRC64;  
  
 Query Match 17.4%; Score 435.5; DB 10; Length 514;  
 Best Local Similarity 29.8%; Pred. No. 1.9e-25;  
 Matches 130; Conservative 79; Mismatches 162; Indels 65; Gaps 17;  
  
 QY 51 VENIWLILENRSDNIGVRRROGLDNPIN--NGPFCYKNASDPSSGKYCTQAKDYDS 108  
 DB 27 IKTVVYVWNRSDHMLGMKMK--LNPEINGVDSNPVSDPSSRKTKFGSGSH-Y 83  
 QY 109 VFNDPDHSTGNNLEFYGTY-----TPNNGAIGASGVVADQSGFLNAQLNDYPKLAPEE 162  
 DB 84 VDPDGHSGFAIREQVFGSNDTSMDDPPMNGFVQQA-YSEDPFGNWSA----- 130  
 QY 163 ATQWNGYTYEEVPTLVLDVDEFTTNSWFCVPGTNPRLCALAGTAGHKNDDDF 222  
 DB 131 ---SVWNGFEDKVPVYKSLVSEFAVDFORFASVPSTQPNRMFVHSGTSAGATSNPIS 187  
 QY 223 LNTGSSKSIPEANKEGWSLNYDGTNGEPEPDSLFYFVYNOT-----SRSNVVPVE 275  
 DB 188 LAKGYQRTIFDNLDDSEFSF---GIYQNPVAVLFYQSLRKLKYVFKPSYGN----- 238  
 QY 276 NFQDAYLGLVLPKSYINPSCCGT---NFTNSMHTGNVSGVEFVKQIYDAIROGPQWDK 332  
 DB 239 SFKDHAKQKLPATVIEQRYMDTLLEPASDDHSHDVFQOGKFKEVYETLRASPOWNE 298  
 QY 333 TLLFITTYDGTGGFDHVPPL--AVRPDNLTYTETAKNGQKTYTLHFDRLGRMPTWVISP 390  
 DB 299 TLLIITYDEHGDFDHPVTPVRNVPSDGIIV-----GPDPELFQFNRLGIRVPTIAVSP 352  
 QY 391 YSKKGYIEQYGTDPVTKAPAP---YSATSVLTKLGLYLWIDED--FTPRVHSPSFDHL-- 443  
 DB 353 WIEKGVV-VHGNP---GSPPPSEYEHSSIPATVKKLFNLSSPFLTKRDEWAGTFENILQ 408  
 QY 444 IGTTLREDAPIALKTP 459  
 DB 409 IRKEPTDCPTLPEP 424  
  
 RESULT 7  
 ID P95245 PRELIMINARY; PRT; 508 AA.  
 AC P95245;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DE PLCC  
 GN PLCC OR RV2349C OR MTCY98.18C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
  
 OC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RA MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Radcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Feltham K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter K., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 DR EMBL: Z83860; CAB06146.1; -.  
 DR TUBERCULIST: RV2349C; -.  
 SQ SEQUENCE 508 AA; 55047 MW; A0A517ABD20BBAA2 CRC64;  
  
 Query Match 13.0%; Score 324.5; DB 2; Length 508;  
 Best Local Similarity 26.0%; Pred. No. 6.6e-17;  
 Matches 140; Conservative 61; Mismatches 203; Indels 135; Gaps 24;  
  
 QY 10 LGLASLGLASPVTSVTSVREAPFGYKPGSKESIEIENLKDKVENIVWLILENSFDNIIG 69  
 DB 7 LAKAGAGAAAVLTDWAAPIEKAAGAGPCS---CHLTD-IEHIVLCLENRSFDFHG 61  
 QY 70 GYVR--QGLDNPINNGPFCYKNASDPSSGKYCTQAKDYDSV-----F 110  
 DB 62 TLSAVDGGFDTP---PLFOQKGNPE-----TQALDPTGITLPRVINTTGGPVGECV 112  
 QY 111 NPDHSTVGNLNFYGTYPNNGATASGVVADQSGFLNAQLNDYPKLAPEEATROVMGY 170  
 DB 113 NPDHSTVGNLNFYGTYPNNGATASGVVADQSGFLNAQLNDYPKLAPEEATROVMGY 153  
 QY 171 YTEEVPTLVLDVDEFTTNSWFCVPGTNPRLCALAGTAGHGN-KNDDDFLNTYGIS 229  
 DB 154 YARPDIHYLLADTTICDQYFSLGGLTNPRLYISATVNPDPGQGGPQIPEAIOQ 213  
 QY 230 KSI-----EANEKGVSW-----LNYDGTNGEF---EPDSLFYFVY 264  
 DB 214 KLTFTWRTMPQNLSDAGISKMYNSKLLGLNDTSLSRNGYVSGFQKQADPRS----- 266  
 QY 265 QTSRNVVPVE--NFQDAYLGLVLPKSYINPSCCGTNTNSMHTGNVSGVEFVKQIYD 322  
 DB 267 DLARYGIAPYPMDFIRDVINNTLFOVSWVVP---LTVESEHPSFPVAVGAVTIVNLIR 322  
 QY 323 AIROGPO-WDKTLLIFTYDGTGGFDHVPPLAVRPDNLTYTETAKNGQKTYTLHFDOR--- 378  
 DB 323 VLLRNPVWERTALLIAYDEHGDFDHPVTPVLA--PEG-TPGEWIPN---SVDIDKVDG 375  
 QY 379 -----LGRMPTWVTSYKGYIEQYGTDPVT-----GKPAYSATSVLTK 420  
 DB 376 SGGIRGPTGLGFRVPCFVSPYSGGLMVHDFDHTSOLQLIGKRFGRVYPNPTWPRASV 435  
 QY 421 LGYLWIDIEDF--TPRVAHSPSFDH-----LIGTTLREDAPIALKTPHTFSV 464  
 DB 436 TGDMTSAFNAFAPDPSPNLDHPVRLQPKVAKCPVNPVLGFLNGLGFLYRVPYPTTPV 494  
  
 RESULT 8  
 ID Q9RGS8 PRELIMINARY; PRT; 700 AA.  
 AC Q9RGS8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DE PLCC  
 GN NON-HEMOLYTIC PHOSPHOLIPASE C.  
 OS Burkholderia pseudomallei (pseudomonas pseudomallei).  
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;

OC Burkholderia  
 OX NCBI\_TaxID=28450;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPT1;  
 RX MEDLINE=99454884; PubMed=10523590;  
 RA Korbsrisate S., Suwanasai N., Leelaporn A., Ezaki T., Kawamura Y.,  
 RA Sarasombath S.;  
 RT "Cloning and characterization of a nonhemolytic phospholipase C gene  
 from Burkholderia pseudomallei";  
 RL J. Clin. Microbiol. 37:3742-3745(1999).  
 DR EMBL: AF107252; AAF17299.1; -.  
 SQ SEQUENCE 700 AA; 77190 MW; 5EAD6GFC0DFB129B CRC64;

Query Match 12.7% Score 318; DB 2; Length 700;  
 Best Local Similarity 27.8% Pred. No. 3.3e-16;  
 Matches 126; Conservative 55; Mismatches 160; Indels 112; Gaps 24;

QY 6 LLTGIGLLASGLASPVTSYTSVREA---PFGYKPGSKESIEIENLKDKNVENIWLILNLR 62  
 DB 12 LAAGTAGAALOLFPPV-----IREALAI PANRRGT-----IRD-VEHIVILMQENR 58  
 QY 63 SFONILG---GVRROGLDNP---NNGPFCNKNASDPSSGKYCTQAKDYDSVFNDDPHS 116  
 DB 59 SFDHYFGKLRGVGFGDPRPLTLONGKPVHPQVLLGPA-----ELLPHFPDAS 107  
 QY 117 VTGNLEF-----YGTYPNNGAIGAGKVVADQSGFLNAQLNDYPKLAPEEATROVMGY 171  
 DB 108 NLG---MQFLQDLPHG---WQDTHGAWNKG---YDRWIANKGT-TTMAYL 148  
 QY 172 TEERVPTLDVDEFTTNSWFCVPGPTNPRLCALAGTAAGHGKNDDEL----- 226  
 DB 149 ERGDIPHYQLADATFICDAYHCSIPSDPNRYMYWTGYVNGDAGGCVLGNREAGY 208  
 QY 227 ISSKSIPEAANEKGVSWLND---GTN-----GFEEDPSLFF----- 260  
 DB 209 WSYTP---ETLEAGVSWKIYODITGLDAGSWGWTNPYIGNYGDNALLYFNQYRNAQP 266  
 QY 261 ---TYVNOTSRNVYVVENF---QDAYLGLVPKFSYINPSCGNTNMSHPTGNVSYG 313  
 DB 267 GSPLYDKARTGTVNSAGTGLFDVLOQDVKNGLTPOVSWI---CAPEAYSEHPNWPANYG 322  
 QY 314 EVFKQIYDAIRGQFQ-WDKTLFIYDTFGFDYHVPPLA--VRPDMLTYYTETAKNGQ 370  
 DB 323 AMVVEQVLTALTSNPDVWSKTFALFYDENGFFDHFVAPPAPQSRDGLSTVSTA--GE 380  
 QY 371 KY---TLHFD---RLGGRMPTWISPSYKGYI 397  
 DB 381 IFPGDAAHMAGPYGLGPRVPMVVPWTKGGWV 413

RESULT 9  
 Q9XB13 PRELIMINARY; PRT; 514 AA.  
 AC Q9XB13;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
 DE PHOSPHOLIPASE.  
 GN RV02-PLCDB.  
 OS Mycobacterium bovis BCG.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=33892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BCG-PASTEUR;  
 RX MEDLINE=99255698; PubMed=10320585;  
 RA Gordon S.V., Brosch R., Billault A., Garnier T., Eiglmeyer K.,  
 RA Cole S.T.;  
 RT "Identification of variable regions in the genomes of tubercle bacilli  
 using bacterial artificial chromosome arrays.";

RL Mol. Microbiol. 32:643-655(1999).  
 DR EMBL: Y18606; CAB44656.1; -.  
 SQ SEQUENCE 514 AA; 55770 MW; 9D99A46D312BA30 CRC64;

Query Match 12.3% Score 308; DB 2; Length 514;  
 Best Local Similarity 28.0% Pred. No. 1.2e-15;  
 Matches 124; Conservative 45; Mismatches 156; Indels 118; Gaps 20;

QY 9 GLGLASGLASGLASPVTSYTSVREA PFGYKPGSKESIEIENLKDKNVENIWLILNRSFDNIL 68  
 DB 22 GAGALMS--FAGPVI-----EKAYGAGPCS-----GHLTD-IEHEVFFMQENRSFDHYF 67  
 QY 69 GGVRRROGLDNPINNGPFCNKNASDPSSGKYCTQAKDYDSV-----F 110  
 DB 68 GTL---SCTDGFNTVSFLFQKGNP-----MTQALDATGVTWMPYRDTTRGPFLLDGACV 119  
 QY 111 NDPHSVTGNLEFYGTYTPNNGAIGAGKVVADQSGFLNAQLNDYPKLAPEEATROVMGY 170  
 DB 120 NDPHS-----WVAMHESWNGCV-----NDNWLPAQ---AKTRSAAHPTVMGY 160  
 QY 171 YTEEVPTLDVDEFTTNSWFCVPGPTNPRLCALAGTAAGHGKNDDEL----- 223  
 DB 161 YTRDIPHIYLLADAFVCDYFCVSLGPTLPRILYLSATIDPDGONGGPELQSPFTQP 220  
 QY 224 ---NYGISKSIPEAANEKGVSWLNDGTNGEFEEDPSLFTYYVNOTS----- 267  
 DB 221 VRRGWRIMPONLSA---GVSWKVY---RNKTLGPISSVLTYSGLVTSFKQSDPRSDL 274  
 QY 268 -RSNVVP---VENFQDADYGLVPKFSYINPSCGNTNMSHPTGNVSYGEVFKQIYDAI 324  
 DB 275 VRFCVAPSPASFAADVLNRLPRVSWIPNV---LESEHPVAPAAAGAFVILRL 330  
 QY 325 RQGFQ-WDKTLFIYDTFGFDYHVPPLA VRPDMLTYYTETAKNGKTYL-HFDR----- 378  
 DB 331 LANPAVWEKTALVSYDENGFFDHFVVPATA-----PAGTPGEYTVTPDIDQVPGS 381  
 QY 379 -----LCGRMPTWISPSYK 393  
 DB 382 GGIRGPIGLGFRVPCFVISPYSR 404

RESULT 10  
 P95246 PRELIMINARY; PRT; 512 AA.  
 AC P95246;  
 DT 01-MAY-1997 (TReMBLrel. 03, Created)  
 DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
 DE PLCB.  
 GN PLCB OR RV2350C OR MTCY98.19C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 DR EMBL: Z83860; CAB06147.1; -.  
 DR TUBERCULIST; RV2350c; -.  
 SQ SEQUENCE 512 AA; 55623 MW; B82672FA0C81DA9B CRC64;

```

Query Match      11.5%; Score 288; DB 2; Length 512;
Best Local Similarity 26.1%; Pred. No. 4.3e-14;
Matches 141; Conservative 60; Mismatches 184; Indels 156; Gaps 29;

QY 14 ASGLASVPTSEYTSVREAPFGYKPGKSEIENLKDKNVNIWLINRSDNGLGGVR- 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 AFMSLAGPIIEKAYGA-----GPCPG-----HLTD-IEHIVLLMQENRSDFHYFTLSD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 73 RGLOD-----NPIINNGPFCNKYKASDSSGKYCTQAKDYDSV-----F 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 TRGFDDTTPVVAQSQNPMI-----QAVDPAG---VTLPRDFTTGRGLVAGECV 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 111 NPDHSVTVGNLLEFYGTYPNNGAIASGVVADQSGFLNAQLNDYPKLAPEEATRVMGY 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 NPDHSHWIGMNSW-----NGGA-----NDNWLPAQVPFSPLOQGNVPVT--MGF 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 171 YTEEVPTLDVLDEFTFNSWFCVPGPTNPNRLCALA-----GTAAGH--GKNDD 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 YTRDLPPIHYLLADTFTVCDGYFCSLLGGTTPNRLYMSAWIDPDGTDGPGVLIENIQP 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 222 FLNYGISKSIIFAEANEKGVSW-----LNVDGTNGEF-----EPDSLFTTY 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 LQHY--SWRIMPENLEDAGVSKYQNKLLGALNNTVVGYNGLVNDKQAADPRS----- 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 263 VNQTSRSNVVPE--NFFQDAYLGLVLPKFSVINPSCCGTNTNSMHPGTGVSVEVFKQI 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 --NLARFGISPTPLDFAADVRNNRLPKVSWVLPGF---LLSEHPAPFVNVGAV--AI 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 321 YDAIR--QGQW-WDKTLFIITYDTGFDVHVP--LAVRPNLTYYTETAK 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 VDALKILLNSPAVWEKTAALIVNDENGCFDHFVVPPTPPGPGFVTVPD--IDSVPGS 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 368 NQOKYTLHFDRLGRMPTWVISPYSKKGYEIOXGTDPT-----GKPAPYSATSV 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 GGIRGPI--GLGRFVPCLVISPYS-RCPLMVHDTFTDHTSTLKLIRARFGVVPNLTAWR 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 418 LKTLGLYMDIEDP-TPRVAHSPSFDH-----LIGTTLREDADIALKTPH 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 DATVGDMTSTFNEAAPNPSPKPLNDHPRLNALPKLPQCVPNNAVILGTVTAIRYRVPFQ 494
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 461 T 461
   :
Db 495 S 495

RESULT 11
Q50360
ID Q50360 PRELIMINARY; PRT; 520 AA.
AC Q50560; Q50771; O08223;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PHOSPHOLIPASE C.
GN MPCA OR PLCA OR RV2351C ORMTCY98.20C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Johansen K.A.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE OF 9-520 FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=91372944; PubMed=1909999;
RA Parra C.A., Londono L.P., del Portillo P., Patarroyo M.E.;
RT "Isolation, characterization, and molecular cloning of a specific
   Mycobacterium tuberculosis antigen gene: identification of a species-
   specific sequence.";

```

```

RN Infect. Immun. 59:3411-3417(1991).
[3]
RN SEQUENCE OF 9-520 FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=92042637; PubMed=1939567;
RA del Portillo P., Murillo L.A., Patarroyo M.E.;
RT "Amplification of a species-specific DNA fragment of Mycobacterium
   tuberculosis and its possible use in diagnosis.";
RL J. Clin. Microbiol. 29:2163-2168(1991).
[4]
RN SEQUENCE OF 9-520 FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
   Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
   Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
   Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
   complete genome sequence.";
RT Nature 393:537-544(1998).
RL EMBL; U49511; AAC18943.1; -.
DR EMBL; L11868; AAB59164.1; -.
DR EMBL; Z83860; CAB06148.1; -.
DR TUBERCULIST; RV2351C; -.
KW Hypothetical protein.
SQ SEQUENCE 520 AA; 56138 MW; 5844EFOC064288A7 CRC64;

Query Match      11.4%; Score 286; DB 2; Length 520;
Best Local Similarity 28.6%; Pred. No. 6.2e-14;
Matches 128; Conservative 40; Mismatches 161; Indels 118; Gaps 20;

QY 7 LTGLGLASL-GLASVPTSEYTSVREAPFGYKPGKSEIENLKDKNVNIWLINRSD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 LTGAGAAFLMDWAAPVIEKAYGA-----GPCPG-----HLTD-IEHIVLLMQENRSFD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 66 NILGVRVRLQGLDNPINNPFCKYKASDSSGKY-----CTQAKDYDSV----- 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 HYFGTLSS-----TNGFNAASPAFQGMNPMPTQALDPAQVTIPFLDTRGP 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 110 -----FNDPDHSVTVGNLLEFYGTYPNNGAIASGVVADQSGFLNAQLNDYPKLAPEEA 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 FLDGECVNDPEHQWGMHLAW-----NGGA-----ND--NWLPAQA 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 164 TRQ-----VMGYTTEEVPTLDVLDEFTFNSWFCVPGPTNPNRLCAL-AG 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 TTRAGPVPLTNGYTYTRQDIFIHYLLADTTTICDGYHCSLLTGLTPNRLYNSANIDPAG 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 211 TAAGHGKNDNDFLNYGISKSIIF-EAANEKGVSWLNVDGTN-GEFEPDSLFTTYVNOTSR 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 TDGGPQLVEPGFLPLQOQFSWRIMPENLEDAGVSKYQNKGLGRFINTPISNNGLVQAFR 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 269 SNVYVPEV-----FFQDAYLGLVLPKFSVINPSCCGTNTNSMHPGTGVSVEY 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 QADPRNLARYGIAPTPGDFDAADVRANRLPKVSWLVPNT-----LOSEHPALVALGAV 323
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 316 FVKQIYDAIRGQW-WDKTLFIITYDTGFDVHVPPLA-----VRPNLTYYTETA 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 SMVTAIRLLNSPAVWEKTAALIVSYDENGCFDHFVTPPTAPPPTGPGFVTVPEVNIID-AVPG 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 367 KNGOKYTLHFDRLGRMPTWVISPYSK 393
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 SGGIRGPI---GLGRFVPCLVISPSYR 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 12
Q50561
ID Q50561 PRELIMINARY; PRT: 521 AA.
AC Q50561;
DT 01-NOV-1996 (TremBrel. 01, Created)
DT 01-NOV-1996 (TremBrel. 01, Last sequence update)
DT 01-NOV-1996 (TremBrel. 01, Last annotation update)
DE PHOSPHOLIPASE C.
GN MPCB.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Johansen K.A.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U49511; AAC18944.1; -.
SQ SEQUENCE 521 AA; 56424 MW; 7C915196E6914256 CRC64;

Query Match 10.8%; Score 271; DB 2; Length 521;
Best Local Similarity 25.5%; Pred. No. 8,9e-13;
Matches 138; Conservative 60; Mismatches 186; Indels 158; Gaps 29;

QY 14 ASGLASPTVSEYTSVREAPFGYKPGSKESIEENLKDKVENIWLLENRSFDNIGVGR- 72
DB 27 AFMSLAGPIIEKAYGA-----GPCPG-----HLTD-IEHIVLLMOENSRDHFYGLTSD 74
QY 73 ROGILD-----NPNNGPCNYKNASDPSSGKYCTQAKDYDSV-----F 110
DB 75 TRGFDDTTPPVVFAQSGNPNWT-----QAVDPAG---VTLPYREDTRGPLVAGECV 123
QY 111 NDPDHSVTNNLEFYGYTTPNCAIASGKYVADSGFLNAQLNDLPKLAPEATQVMGY 170
DB 124 NDPDHSNIGMHSN-----NGCA-----NDNWLPAQVPFSPLOGNPVPT---MGF 165
QY 171 YTEEEVPTLVLDVDEFTTNSWFSVCPGPTNPNRLCALA-----GTAAGH---GKNDSD 221
DB 166 YTRRDLPPIHYLLADFTVCDGYFCSLLGTTNRLYNSAWIDPDGTDGCPVLIEPNIQ 225
QY 222 FLNYGSSAKIFEANKEGVSW-----LNVDGTNGEF-----EPDSLEFFY 262
DB 226 LQHY--SWRIMPENLEAGVSWKYQNKLLGALANTVVGYNGLVNDFOAADPKRS----- 278
QY 263 VNQTSRSNYPVPE--NFFODAYILGVLPKFSYINPSCGTTNTSMHPTGNVSYGVFVKQI 320
DB 279 --NLARFGISPTYPLDFAADVARNRLTKVSWVLPGF---LLSEHPAPPVNVGAV---AI 329
QY 321 YDAIR---QGPO-WDKTLLETTYDETGFGYDHVPPP-----LAVRPDNLITYTETAK 367
DB 330 VDAILRLLSNPAVWEKTAIVNYDENGDFDHFVPPPTPPGTPGEFVVPD--IDSVPGS 387
QY 368 NGQKYLTHDRLGGRMPTWVISPYSKGYIEQYCTDPVT-----GKPAPYSATSV 417
DB 388 GGIRGAI---GLGPRVPCLEISPYT--GPLMVHDTFDHTSTFLKIRAREGVVPLNTAWR 442
QY 418 LKTLGLVNDIEDF-TPRVAHSPSFDH-----LIGTTLREDAPIALKTP 459
DB 443 DATVGDITSTFNAAPPNPNKPLNDHPRAQCALKPLQCPVNAVLGTVTKTAIPRYVPFP 502
QY 460 HT 461
DB 503 QS 504

RESULT 13
Q9L5C5
ID Q9L5C5 PRELIMINARY; PRT: 577 AA.
AC Q9L5C5;
DT 01-OCT-2000 (TremBrel. 15, Created)

```

```

DT 01-OCT-2000 (TremBrel. 15, Last sequence update)
DT 01-OCT-2000 (TremBrel. 15, Last annotation update)
DE ACID PHOSPHATASE.
GN ACPA.
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=28450;
RN [1]
RP SEQUENCE FROM N.A.
RA Burtick M.N., Bolton A.J., Brett P.J., Watanabe D., Woods D.E.;
RL Identification of acpA gene homologs in pathogenic and non-pathogenic
RL Burkholderia species facilitates TnpA mutagenesis.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF252862; AAF6061.1; -.
SQ SEQUENCE 577 AA; 62839 MW; 16DAAD567D973D82 CRC64;

Query Match 8.2%; Score 206; DB 2; Length 577;
Best Local Similarity 21.6%; Pred. No. 1e-07;
Matches 114; Conservative 69; Mismatches 176; Indels 170; Gaps 23;

QY 10 LGLSLASGLASPTVSEYTSVREAPFGYKPGSKESIEENLKDKVENIWLLENRSFDNIL 68
DB 45 LGGIALGASVALAGCETAVPAAP---RSMALRLDITLARREVRNIVVIAENSFANLF 101
QY 69 GGYRROGLDNPINNGPCNYKNASDPSSGK-----YCTQAKDYDSV-FNDPDIH 115
DB 102 GDF--PGVQOPL--GAVSPERYAQLDRDGKTLATLPKIMGVLVPOQAVDQKRYMIGE 157
QY 116 SVTG--NNLEF-----YGYTTPNNGAIASGVVAD-----QSGFLNAQLNDYKLAPEE 162
DB 158 DIVGLNPAFFVTPDAQCKPLPN-----GVITRLWHREYQNMQINGGRND-QFAAWAD 210
QY 163 ATQVMGYI-TEEEVPTLVLDVDEFTTNSWFSVCPG-----PTNPN----- 203
DB 211 SGLVLMGHRNSAETLRLNLARGYTFCDNFMAAFGGSWLNHIFLISAQAPLYPDVHQ 270
QY 204 -----RLCALAGTAA-----GHGKNDDDEL-----NYGISS----- 229
DB 271 PAKHLVVVEGDDPAGTRLKLAADSPASALDGPFPKFNEDGLFTPDGAVNTMAPPYQSS 330
QY 230 -----KSIPEANEKGVSWLNVDGT----- 249
DB 331 VRPVEGNPAYADSNRPVLPQHYATIGDLSEKGVDMWYGANQYALEHRTGSPVD 390
QY 250 -NGEFPDLSLFFTYVNOTSR-----SNVVPVENFTQDAYLGLVPKFSYINPSCCG 298
DB 391 FOYHHQPFNYFASYPAGTEARRRHLRDAGLDGDDPSTNHLIADIDAGRLPTVTFYKPO--- 447
QY 299 TINTSMHPTGNVSYGVFVKQIYDAITROGPQWDKTLFFITYDGTGGFYDHPPLAVRPD 358
DB 448 GNLMMIAGVADVASGDRIHTVIEHMRGPQWHAHTVLLTVDENGGWDPVSP----- 501
QY 359 NLTYTETAKNGOKYTLHFRDLGGRMPTWVISPYSKGYIEQ---YGTDPV 405
DB 502 -----KGDW-----GPGSRIPAMVISPFAKKGVDHTLYDTNSI 536

RESULT 14
Q47936
ID Q47936 PRELIMINARY; PRT: 514 AA.
AC Q47936;
DT 01-NOV-1996 (TremBrel. 01, Created)
DT 01-NOV-1996 (TremBrel. 01, Last sequence update)
DT 01-NOV-1996 (TremBrel. 08, Last annotation update)
DE ACID PHOSPHATASE PRECURSOR.
OS Francisella tularensis.
OC Bacteria; Proteobacteria; gamma subdivision; Francisella group;
OC Francisella.
OX NCBI_TaxID=263;
RN [1]
RP SEQUENCE FROM N.A.

```

[illegible]

RESULT	15
Q9XAN9	
ID	PRELIMINARY; PRT; 501 AA.
AC	Q9XAN9;
DC	01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DD	01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE	PUTATIVE PHOSPHOLIPASE C (FRAGMENT).
SC	SC4C6.01.
OS	Streptomyces coelicolor.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
NB	NCBI_TaxID=1902;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	STRAIN=A3(2);
RC	Sieger S., Harris D.;
RA	"A set of ordered cosmids and a detailed genetic and physical map for
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RT	

PL	Submitted (JUN-1999) to the EMBL/GenBank/DBDJ databases.
RM	[2]
RC	SEQUENCE FROM N.A.
RA	STRAIN=A3(2);
RC	James K.D., Parkhill J., Bartell B.G., Rajandream M.A.;
RM	Submitted (JUN-1999) to the EMBL/GenBank/DBDJ databases.
RL	[3]
RC	SEQUENCE FROM N.A.
RA	STRAIN=A3(2);
RC	MEDLINE=97000351; PubMed=8843436;
RM	Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cul-
RA	Kinashi H., Hopwood D.A.;
RC	"A set of ordered cosmids and a detailed genetic and phy-
RM	the 8 Mb streptomycetes coelicolor A3(2) chromosome.";
RL	Mol. Microbiol. 21:77-96(1996).
DR	EMBL; AL079355; CAB45565.1; -;
FR	NON_TER 1
FT	NON_TER 501
SQ	SEQUENCE 501 AA; 54844 MW; COD7934446F0F369D CRC64;

Query Match 8.1%; Score 202.5; DB 2; Length 501;  
Best Local Similarity 26.7%; Pred. No. 1.6e-07;  
Matches 63; Conservative 28; Mismatches 78; Indels 67; Gaps 8;

QY	210	GTAAAGCKNDDDFLNYGISSKSFIFEAANEKGYSWLNY-----DGTNGE	252
		:  :	
Db	7	GPVLGNDELGYDWTYP-----ERLEAAGYSMKIYQDVGDLDAAGHGWIDDAYRGN	59
		:  :	
QY	253	FEPOSLEF-----TYVNOTSRNVVPVENFQO---DAYLGVLPKFSYINPSC	296
		:  :	
Db	60	YGDNSLLYFNKYNNAKPGDPLYDKARTGDARKGEGYFEDRLADRVAKGLPEISWI----	115
		:  :	
QY	297	CCTNTSMHPGTNGVSGVEFVKQIIDAIRGPO-WDKTLFTFYDBTGFDYHVPPLAV	355
		:  :	
Db	116	RAPEAFSEHSNWPNSYGAWISQVLDALTSNPKWAKTALFTYDENGDFDHHVVPPLP-	174
		:  :	
QY	356	RPDNLTYTETAKNGO-----KYTLHFDRGGRMFTWVISPYSKKGYI	397
		:  :	
Db	175	-----PKSAQAQGSTVDVSLDFVFGSAAKHREGFYGLGPRVPMLVPSWQSGGFV	223

Search completed: February 26, 2001, 19:32:37  
Job time: 11518 sec



10

11

12

13